REPORT RESUMES

ED 003 607 56

INTERPOLATED INFORMATION AND STUDENT RESPONSE FEEDBACK DURING "BREAKS" IN TELEVISED AND RELATED MODES OF INSTRUCTION.

BY- GAMES, PAUL A. AND OTHERS
OHIO UNIV., ATHENS
REPORT NUMBER NDEA-VIIA-863 PUB DATE 63
GRANT OEG-7-45-0440-169
EDRS PRICE MF-$0.27 HC-$6.44 161P.

DESCRIPTORS- *TELEVISED INSTRUCTION, TELEVISION, *PROGRAMED INSTRUCTION, *FILMS, SOUND FILMS, *COLLEGE STUDENTS, *STUDENT REACTION, STUDENT BEHAVIOR, PERFORMANCE, ACADEMIC PERFORMANCE, FEEDBACK, ATHENS, OHIO

THIS STUDY ATTEMPTED TO EXAMINE THE POSSIBILITY OF INCORPORATING CERTAIN CHARACTERISTICS OF PROGRAMED INSTRUCTION INTO LECTURES PRESENTED THROUGH TELEVISION OR RELATED MODES OF INSTRUCTION. SUBJECT AREAS PROGRAMED WERE SPEECH AND STATISTICS. THE FIRST STUDIES WERE DIRECTED TOWARD THE ROLE OF "BREAKS" IN PRESENTATIONS THAT PERMITTED STUDENT RESPONSES TO MULTIPLE-CHOICE QUESTIONS. LATER STUDIES WERE CONCERNED WITH THE ROLE OF BEHAVIORAL OBJECTIVES AS INTERPOLATED MATERIAL DURING "BREAKS." THE FIRST EXPERIMENTS CONTRASTED TELEVISION PRESENTATIONS IN WHICH INTERPOLATED MATERIAL CONSISTED OF EITHER (1) STEM AND ANSWER OF A "TYPICAL QUESTION," OR (2) A MULTIPLE-CHOICE QUESTION. NO SIGNIFICANT DIFFERENCES WERE FOUND. LATER EXPERIMENTS CONCENTRATED ON VARIATIONS IN THE CONDITIONS OF PRESENTATION OF THE BEHAVIORAL OBJECTIVES DURING "BREAKS." THE USE OF PRINTED SHEETS OF THE OBJECTIVES PRODUCED HIGHER MEANS THAN WHEN THE OBJECTIVES WERE ABSENT. OTHER EXPERIMENTS OFFERED NO CLEAR EVIDENCE OF THE SUPERIORITY OF "BREAKS WITH VISUAL OBJECTIVES" OVER "DEAD AIR BREAKS." HOWEVER, THESE TWO CONDITIONS WERE SUPERIOR TO CONTINUOUS PRESENTATIONS ON "OPEN-END" TESTS. SUGGESTIONS ARE MADE FOR POSSIBLE EXPERIMENTS TO CLARIFY THE DISCREPANCIES ASSOCIATED WITH A NUMBER OF THE FINDINGS. (JC)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

INTERPOLATED INFORMATION AND STUDENT RESPONSE FEEDBACK DURING "BREAKS" IN TELEVISED AND RELATED MODES OF INSTRUCTION

Paul A. Games and F. Craig Johnson, Principal Investigators
with the assistance of
Kenneth Frandsen, Presley Holmes, George Klare, and Patrick Lucas

Title VII Project Number ________
National Defense Education Act of 1958
Grant Number 7 - 45 - 0440 - 169

Ohio University
Athens, Ohio

The Research Reported Herein Was
Supported by a Grant from the
U. S. Department of Health, Education, and Welfare
Office of Education

ED003607

Abstract

INTERPOLATED INFORMATION AND STUDENT RESPONSE FEEDBACK DURING "BREAKS" IN TELEVISED AND RELATED MODES OF INSTRUCTION

Purpose

The studies described here were primarily attempts to incorporate some of the characteristics of programed instruction into group instruction procedures. The first studies were chiefly directed toward the role of various "breaks" in program presentations that permitted overt student responses to interspersed multiple-choice questions with or without feedback to the instructor. Later studies were directed toward the role of "behavioral objectives" as interpolated material during breaks in a performance; the forms used provided for their use as overt and/or covert responses and as introductory material.

Procedure and Results

Behavioral objectives were written for three Speech and four Statistics contents. Test items were constructed for each objective and were used in tests on the various contents. Scores on these tests constituted the criterion measures in the several experiments conducted.

The first experiments contrasted televised presentations in which the pertinent interpolated material consisted of either 1) the stem and answer of a "typical question" combined into a single statement, or 2) a complete multiple-choice question. Students responded (1) on paper or (2) by the use of a numbered switch; student answers in some studies were communicated to the instructor by a mechanical device. In each case, the interpolated material was on the content just covered in the presentation. Neither the experiment on the Speech contents, nor the experiment on the eight televised presentations of the Statistics contents, yielded any significant differences in achievement following these three conditions.

Later experiments concentrated on variations in the conditions of presentation of the actual behavioral objectives during the breaks instead of the "typical questions." Speech experiments involved presentations on audio tape, and the Statistics experiments involved filmed presentations. A Speech experiment failed to show any effect associated with the use of a different voice instead of the regular speaker's voice in reading the interpolated objectives, nor did it show any beneficial effects associated with the use of a printed sheet that required the subjects to fill in the key words of

the objectives that were read. However, the use of printed sheets of the complete objectives produced significantly higher means than when these printed objectives were absent. The greatest superiority occurred when the printed objectives were combined with "dead air" breaks rather than breaks in which the objectives were read aloud by a speaker. This finding occurred not only for the three content replications in this experiment, but also in a later replication involving printed objectives with key words capitalized. The later Speech experimentation failed to yield any differences between the use of the complete printed objectives, the complete printed objectives with key words capitalized, and printed sets of key words alone. The key words were short phrases or words that were judged to be the crucial "paired-associates" in many of the Speech objectives. Unfortunately, the large error term of the last Speech comparisons make highly dubious any conclusion that paired associate "key words" may be substituted for complete objectives.

The last Statistics experiment involved films as teaching devices and contrasted three introductory periods involving (1) a pretest of one item for each objective, (2) study of the objectives themselves, or (3) study of extraneous material. A second dimension was added by contrasting presentations using interpolation of the behavioral objectives during breaks by means of slides vs. "dead air" breaks vs. the continuous film with no breaks. This design was replicated on two subject matters, with different groups of subjects taking different sequences.

With these contents and the filmed presentation, there was no clear evidence of the superiority of the "breaks with visual objectives" over the "dead air breaks." However, these two conditions were superior to the continuous presentation condition on "open-end" tests. The introductory groups with the pretest or study of the objectives were found to be significantly superior to the extraneous material group on the Frequency Distribution open-end test. A retention session on this same test two weeks later showed a similar superiority only for the group that had had the continuous presentation. It seemed possible that such aids have long-term benefits only when the conditions of original learning are rather inefficient.

Positive transfer was predicted from the Frequency Distribution film to the Central Tendency test, but no evidence was found of such transfer. Evidence was found for positive transfer from the Central Tendency film to the Frequency Distribution test. This was attributed to a short review section at the start of the Central Tendency film which covered five of the Frequency Distribution objectives, and to the additional trials on Frequency Distribution concepts that were used at many places during the Central Tendency film. The Frequency Distribution retention tests also showed a clear superiority of subjects who had seen the Central Tendency film during the intervening week over those subjects who had not seen this film.

Suggestions were made for possible experiments to clarify the discrepancy between the superiority of the use of printed behavioral objectives during breaks over the "dead air breaks" as found in the audio-tape Speech experiments, and the lack of such superiority for slide presentations of such objectives in the filmed Statistics experiments. Discussion of the possible paired-associates nature of some academic contents vs. possible hierarchical sequence in other contents concluded the report.

INTRODUCTION

Currently there is widespread interest in American education in new teaching devices and methods. Two of the foci of this interest have been instruction by television and by teaching machine. Both hold some promise of increasing the efficiency of teaching by extending an expert presentation to a larger audience than can be reached by an instructor in a conventional classroom.

Early work in both fields suggested that the device used had its importance, but that the critical factor needing study was the "program." In the case of television, the term as used here refers to the presentation (i.e., to the "program" transmitted to the student). In the case of teaching machines, the term refers to the teaching material as it is organized in a special fashion (i.e., a "program" put on the machine or into a "programed" book studied by the student). Each method, through its particular kind of program, has its particular advantages and disadvantages for teaching. In some ways, these appear complementary. Teaching by machine--or programed instruction, as it is now most commonly called--is sometimes criticized for leaving the instructor out of the teaching situation; televised instruction puts him back in, albeit indirectly. Televised instruction is sometimes criticized for sacrificing the productive interchange between the student and the teacher[1]; programed instruction puts it back in, though also indirectly.

There are circumstances, of course, where one or the other method is especially appropriate, e.g., programed instruction for individual study situations and televised instruction where large groups of learners must be reached at one time. The use of one method, however, has heretofore automatically ruled out the advantages of the other. Could the advantages of the one be incorporated in the other?

Purpose of the Study

It was the purpose of the study to examine the possibility of incorporating certain characteristics of programed instruction into lectures presented via television or related modes of instruction (i.e., via audio tape, film, or kinescope recording). Several prior considerations guided the work.

1. No attempt was to be made to analyze the relative effectiveness of televised versus programed instruction *per se*. The problem of control of relevant variables in such a comparison is almost insurmountable.

2. It was considered desirable that the method used be that of group rather than individual presentation, for reasons of efficiency of instruction. This determined that programed instruction characteristics

---

[1] Skinner, B. F. "Teaching Machines." Science, 128, Oct. 24, 1958.

would have to be incorporated into a televised presentation, rather than vice versa.

3. Many added materials may precede or follow a televised presentation in the attempt to improve what is learned from it. Most research has centered on this before-after approach. Study, therefore, was concentrated chiefly on the question of what could be done during the presentation. This meant that one or more short breaks would need to be provided in the televised programs.

4. It was considered desirable that any suggested changes should not be so drastic that they would not be adaptable to the conventional (usual) manner of televising educational programs.

Background

With the above considerations in mind, it was necessary first to examine the characteristics of programed instruction that might be used in the research. Some notion of the many characteristics already studied can be seen in the review by Schramm.[2] Among those considered especially important have been active responding by the learner, provision of feedback to the learner, and use of behavioral objectives to guide instruction. Each of these can contribute to the desired interchange between student and teacher that is lacking in a conventional televised presentation.

Each of these, also, can be included in a live televised presentation modified by the inclusion of short breaks during the program. Use can be made of such intervals to provide for active responding on the part of the learner as opposed to the passive responding of the viewer, listener, or reader in a conventional televised presentation. This can take the form of incorporating paper sheets or mechanical response devices into the presentation, with the subject required to react to stimuli provided in the lecture or through the supplementary devices themselves. These stimuli will be referred to by the class name "interpolated information."

The interpolated information used can take many forms. A form intended to help the learner to know what is important or to provide other related aids to learning (for organization, for review, etc.) is common. For this study it was felt that provision of two kinds of interpolated information might be particularly helpful to the learner: use of "typical questions" and of behavioral objectives. In the study, learners were given multiple-choice tests to provide evidence of immediate and delayed retention, so it seemed likely that questions typical of (but not identical to) those test items later used in retention tests should be helpful. Such items were therefore used. In addition, "behavioral objectives" (statements of instructional goals) were used as interpolated information on the supposition that these guides used by the television lecturer (or the experimenter) in organizing and writing his presentation should be especially useful to the learner while learning.[3] Several variations were used in presentation: the auditory vs. the printed medium; variations in stress (emphasis) in oral presentation (use of a voice different from the lecturer's for the interpolated information) and in printed presentation[4] (key words in the interpolated statements presented in isolation, or set apart from the other words by all-capital letters); and combinations of the above.

---

[2] Schramm, Wilbur. Programed Instruction, Today and Tomorrow. New York: The Fund for the Advancement of Education, 1962.

The provision of breaks during televised presentation, and the use of certain kinds of interpolated information, also made possible feedback of two types.[5] The first was the sort commonly found in programed instruction: provision of knowledge of the results of the learner's response back to him.[6] This was possible where the interpolated information was in the form of the typical questions mentioned above. In a second type of feedback, the instructor could acquire information from student responses in a manner analogous to the classroom situation. This was possible through the use of mechanical devices which permit an instructor to modify an on-going presentation as well as use such information for the modification of subsequent presentations.[7] It has been found, for example, that students who watched a conventional television lesson that had been revised to eliminate deficiencies revealed by "sample class" responses scored significantly higher on achievement tests than did students who watched the original unmodified television lesson.[8]

---

[3] For a complete description of the development and use of behavioral objectives, see Mager, Robert F. Preparing Objectives for Programed Instruction. San Francisco: Fearon Publishers, 1962.

[4] For studies utilizing typographic stress and other special typographic arrangements, see: Klare, George R., Mabry, James E., and Gustafson, Levarl M. "The Relationship of Patterning (Underlining) to Immediate Retention and to Acceptability of Technical Material." Journal of Applied Psychology, 1955, 39, 40-42; and Klare, George R., Nichols, William H., and Shuford, Emir H. "The Relationship of Typographic Arrangement to the Learning of Technical Training Material." Journal of Applied Psychology, 1957, 41, 41-45.

[5] For a general description of this process, see Johnson, F. Craig, and Klare, George R. "Feedback: Principles and Analogies." Journal of Communication, 1962, 12, 150-159.

[6] For general reference to a number of studies of feedback in programed instruction, see Schramm, op. cit. For a specific reference to the importance of feedback for correct answers in televised instruction, see Gropper, George L. and Lumsdaine, Arthur I. "The Contribution of Sequencing, Pretesting, and Active Student Responses to 'Programed Learning' TV Instruction." Report No. 3, Studies in Televised Instruction, Metropolitan Pittsburgh Educational Television Stations and American Institute for Research, Pittsburgh, Penna., 1961.

Another major variable investigated was study method. Writers on study techniques rather consistently state that a quick "survey" of what is to be learned is helpful in subsequent intensive study.[9] The better the tuitional value of the survey material, the greater its effectiveness is presumed to be; the use of the behavioral objectives mentioned above, or items similar to those in a test to be faced later, appeared to be of potential value and were tried out. Another aspect of study method that has a fairly predictable effect is the retroactive interference of subsequent learning upon previously learned material (sometimes studied as negative transfer[10]). It was re-examined in the learning situation used here. A somewhat less well-defined aspect of study method is the effect of "set to learn." This has been operationally defined in terms of use of a difficult learning task and/or test preceding new learning in order to establish a stronger set to learn as opposed to a typically weaker one.[11] The effect of set has been observed in printed presentation, but has apparently not been tested experimentally with oral presentation.

---

[7] See Johnson, F. Craig. "Feedback in Instructional Television." *Journal of Communication*, 1960, *10*, 140-146; and Johnson, F. Craig, Gares, Paul A., Andersch, Elizabeth G., and Penson, Edward M. "An Investigation of Motion Picture Film and the Program Analyzer Feedback to Improve Television Teacher Training." *Speech Monographs*, *30*, 1963, 373-376 (for greater detail see the Final Report by the same title and by the same authors to the U. S. Office of Education covering a research project carried out under Title VII, the National Defense Education Act, 1958).

[8] Cropper, George L., and Lumsdaine, Arthur I. "Improvement of Televised Instruction Based on Student Responses to Achievement Tests." Report No. 1, Metropolitan Pittsburgh Educational Television Stations and American Institute for Research, Pittsburgh, Penna., 1961.

[9] See Robinson, Francis P. *Effective Study* (Revised Edition). New York: Harper and Brothers, 1961.

[10] Travers, Robert M. W. *Essentials of Learning*. New York: The Macmillan Co., 1963.

One of the problems encountered in research on human learning is that cross-validation is relatively infrequent. Where attempts to remedy this have been made, they are too rarely made across varied contents or subject-matters, so that general applicability is not clearly known. And, of course, it appears to be most difficult of all to demonstrate laboratory-spawned effects in a field setting. In this study, an attempt was made to meet these problems through the use of varied contents within each of two subject-matter areas, speech and statistics. Furthermore, most of the work was done by subjects as part of their regular college courses in these areas.

The Problem Summarized

This study was an attempt to answer the following questions:

1. Will subjects learn more from watching televised instruction when:
   a. there is no responding by subjects during the presentation?
   b. relevant multiple-choice questions are interpolated in the presentation?
   c. modification of the on-going presentation is guided by feedback of subjects' answers to relevant interpolated multiple-choice questions?

2. Will subjects learn more from a recorded (audio) presentation when behavioral objectives are:
   a. not interpolated ("dead air" breaks)?
   b. interpolated orally, "with" versus "without" voice changes for emphasis?
   c. interpolated in printed form, "with" versus "without" capitalization and in key word versus complete statement form?
   d. interpolated in combined oral and printed form?

3. Will subjects learn more from a filmed presentation when:
   a. they have, during an immediately preceding period, taken a pretest, studied behavioral objectives, or studied extraneous materials?

---

[11] Klare, George R. The Measurement of Readability. Ames, Iowa: Iowa State University Press, 1963, pp. 179-180.

b. no breaks or "dead air" breaks or breaks with behavioral objectives are interpolated?
c. another statistics topic has preceded the present one?

As indicated, speech and statistics subject-matter areas were used. In addition, each area was divided into separate contents for experimental work, the speech area into parliamentary procedure, phonetics, and propaganda topics, and the statistics area into frequency distribution, central tendency, variability, and normal curve topics.

METHOD

Experimental Materials

Apparatus. As indicated previously, this study consisted of an investigation of several variables in instruction carried out by means of televised and related modes of presentation. In those parts of the study where immediate feedback to the instructor of student responses was a major variable, a live televised presentation was used. Where feedback was not involved, it was more expedient for reasons of both experimental control and expense to use either audio tapes or 16mm. films instead. With all three modes of presentation, a "professional" (ideal) educational production was attempted (i.e., in terms of lighting, sound, training of instructors, etc.) and commercial equipment and techniques were used (i.e., in audio tape recording, conversion of video tape to film, etc.). Therefore it seemed unnecessary here to describe in detail the common apparatus employed. Only one special piece has been described below, with any necessary detail regarding use of apparatus presented later in the Procedure section.

The one special piece of apparatus used was the "Program Analyzer" (PAR). This feedback device has an individual five-position switch for each of thirty members of a viewing audience. It makes it possible for each viewer (i.e., experimental subject) to indicate his answer to a multiple-choice type of question with up to five possible choices (or, in another modification, his judgment of a particular segment of a television presentation on a five-point scale). This information is fed back to the instructor so that he can see the individual answer numbers (not, however, identified by subject) or the average of the answer numbers of the viewing audience to a given question.

The PAR operates in the following manner. Each of the five positions represents a different amount of electrical current. The current activates two kinds of reporting instruments: counters and lamps. The counters are activated through stepper relays and report the number of switches at a given position at a given moment. This number of switches is also indicated by the number of lamps on the PAR panel that are lighted. These lamps provide a graphic distribution of the judgments, and thus permit an instructor to see the viewer's responses to his presentation. A new set of responses can be recorded and relayed to the panel every 45 seconds.

Experimental Learning Materials. As indicated above, two subject-matter areas were selected for study, speech and statistics. Each will be discussed briefly below.

In the area of speech, three separate contents were selected for use. The first was to have a low initial level of student knowledge; phonetics was chosen for this purpose. The second was to contain information about which students would have considerable knowledge but common misconceptions; propaganda was the choice. The third was to be intermediate in level of knowledge, with parliamentary procedure being chosen. It was felt that this arrangement would provide a certain generality of instructional material across which an effect might be shown. No special attempt was made to develop the above characteristics beyond initial selection of the lessons, however, since it was difficult to get objective evidence for them.

Experimental lessons were prepared for each of these contents on the basis of behavioral objectives (see next section). The lessons were planned for twenty-minute length so that they could be used in regular class meetings in Speech 1, the introductory course at Ohio University. The remainder of the class hour was devoted to administrative details and retention testing. The particular contents chosen either had not been covered in the course prior to experimental use or were not normally a part of the course, so that there could be no prior formal exposure to them. The instructors were 18 randomly selected graduate assistants who had taught or were teaching Speech 1.

In the area of statistics, four contents were used: frequency distribution, central tendency, variability, and normal curve. As in the case of the speech materials, lessons were prepared from the above contents on the basis of behavioral objectives previously prepared (see next section). On each of the four contents, two fifty-minute lessons were planned, rehearsed and taught by the same instructor. They were taught in proper sequence as a part of Psychology 105, an elementary statistics course, in the fall semester, 1962. Because of this use, it was necessary to select a textbook that would fit the needs of the course and yet not interfere with the experimental lessons. A book by Guest was chosen for use.[12] The lessons were taught by three Ohio University faculty members and one graduate student, all of whom had had previous experience in teaching statistics.

---

[12] Guest, Lester. Beginning Statistics. New York: Thomas Y. Crowell, 1957.

Behavioral Objectives. As indicated above, behavioral objectives were developed for use in preparing the lessons from the chosen contents and in later experimentation. These are essentially statements of instructional goals in the form of the behavior desired of the learner after learning.[13] As such, they may potentially form a framework for preparation of test or program frames, for review, for test item preparation, and, as in this study, for guidance of the learning process itself.

For the speech area, a set of objectives was written by staff members for each content on the basis of (a) an analysis of appropriate topics in introductory speech texts, and (b) knowledge gained from the experience of technical experts in teaching appropriate topics. Seventy-five objectives were prepared, 19 for propaganda, 28 for parliamentary procedure, and 28 for phonetics. Written texts embodying these objectives were prepared and given to instructors who individually prepared televised (or audio-taped) lessons (presentations) covering them. (Complete sets of objectives are presented in Appendix A.)

For the statistics area, a set of behavioral objectives for each content was written by staff members on the basis of an analysis of relevant material in beginning statistics texts and experience gained in teaching beginning college courses in statistics. One hundred and twenty-three objectives were used as follows: (1) frequency distributions, 39; (2) central tendency, 37; (3) variability, 30; (4) normal curve, 27. These sets of objectives were used by the statistics instructors in preparing televised lessons covering the given topics. (Complete sets of objectives are presented in Appendix A.)

Retention Tests: 1. General Approach. The effectiveness of each of the experimental variables used was measured through immediate and/or delayed verbal retention tests. These tests were constructed in the following manner.

Since a number of persons were to contribute test items, a statement called "Procedures for Item Writing" was first prepared to standardize the work. It was based largely on the book by Wood.[14]

All items written were of the five-alternative multiple choice type. Insofar as possible, all were constructed so that they could be modified for use as fill-in or short-answer items. They were also classified as "direct" or "transfer" items (i.e., were either based directly upon the lessons or involved transfer of knowledge or skill from non-lesson situations). However, this distinction did not prove useful and was abandoned in the actual construction of the tests, since variations in examples and techniques used by the instructor changed the classification. The work done in each of the two subject-areas is described below. (Copies of the tests are presented in Appendix B.)

---

[13] See Mager, op.cit., for a complete description.

[14] Wood, Dorothy Adkins. Test Construction. Columbus, Ohio: Charles E. Merrill Books, Inc., 1960.

Retention Tests: 2. First Statistics Study. Approximately one thousand statistics test items were prepared to cover the six content areas. This resulted in at least four items for each behavioral objective in each area (and more for most). Almost all of the items were written by the authors, but some were taken from the course item file of one of the authors and some were written by members of an intermediate statistics course. All items were edited by the authors. From this pool of items, 12 tests were constructed, three for each of the areas of frequency distribution, normal curve, central tendency, and variability. There was one test item for each of the objectives in each area, but since the number of objectives covered per area varied somewhat, the total number of items per test varied, as shown in Table 1. Also included in Table 1 are average difficulty values.

Table 1

Numbers of Items and Average Difficulty Values in Three Tests for Each of Four Statistics Content Areas

	Test I		Test II		Test III	
Areas	No. Items	Diff. Level	No. Items	Diff. Level	No. Items	Diff. Level
Freq. Dist.	27	.66	21	.33	30	.69
Nor. Curve	25	.63	16	.70	26	.72
Cent. Tend.	30	.68	29	.64	24	.78
Variability	22	.76	20	.64	24	.73

* Tests I and II were used to measure immediate retention, and Test III delayed retention.

Retention Tests: 3. Speech. Approximately 900 speech test items were prepared to cover the areas. At least three items were available for each objective (some of the earlier items were written for sets of speech objectives subsequently discarded). Most of the items were written by the members of two graduate speech seminars as well as by the authors. All items were edited by the authors.

From this pool, a test for experimental use was devised for each of the content areas covered. On the basis of a tryout with 700 beginning speech students, certain items were used in original form, others in revised form, and still others in rewritten form in three 30-item tests. The sources of the items are shown in Table 2.

Table 2

Sources of Items in Tests for Each of

Three Speech Content Areas

Content	Orig. Form	Revised Form	Rewritten Form	Totals
Phonetics	6	5	19	30
Parliamentary Procedure	14	6	10	30
Propaganda	17	7	6	30
Totals	37	18	35	90

These items were used in further tryout work with students; the final form of each of these 30-item tests had the characteristics shown in Table 3.

Table 3

Reliability, Difficulty, and Discrimination Values for Tests in Each of Three Speech Content Areas

	Reliability Values		Item Analysis Average Values	
Content	Kuder-Richardson*	Split-half	Difficulty**	Discrimination***
Phonetics	.56	.71	.48	37%
Parliamentary Procedure	.59	.23	.61	32%
Propaganda	.59	.67	.60	35%

* Formula No. 20

** Indicates average passing value per item (uncorrected for chance).

*** Indicates average superiority in correct answers of high-scoring group (upper 27%) over low scoring group (lower 27%).

In the above analysis, two of the test items showed a very slight negative discrimination (lower 27% over upper 27%), but they were retained because they were needed to cover a particular objective.

Subjects

As indicated previously, the speech studies and the first statistics study were carried out in appropriate Ohio University courses. In the case of speech, the course was Speech 1, required of all freshmen. The various sections of this course are normally organized into classes of about 15, and this therefore became the sampling unit for this part of the study.

In the case of the first statistics study (Experiment I. C.), the course was Psychology 105, a beginning statistics course which is required of all undergraduate majors in psychology but which is an elective for other students. Forty per cent of the 40 subjects were psychology majors, and the rest were majors in various other academic departments at the University. Most of the subjects were of sophomore or junior standing and all previously had at least a college algebra course. The second statistics study (Experiment IV) used volunteer subjects from Psychology 1 courses, prior to the assignment of the statistics chapter in the text or any lectures on this subject by the instructor.

Procedure and Results

This study consisted of a number of experiments designed to answer the related, but different, questions posed in the Introduction. In addition, certain experiments involved the use of Speech content areas and others the use of statistics content areas. Therefore, the study will be broken down by parts in a manner conducive to ready understanding of the answers to the experimental questions as well as in terms of sequence of operations performed. In line with this, the Procedure and Results sections have been combined and presented in integrated fashion for each of these major parts.

Fall, 1962: Experiment I. A. Speech. This experiment was designed to answer the following questions for the subject-matter area of speech.

Will subjects learn more from watching televised instruction when:

1. there is no responding by subjects during the presentation (regular presentation with only the stem and answer of the corresponding multiple-choice questions shown as a statement during the interpolation period)?
2. they respond to relevant multiple-choice questions interpolated in the presentation (regular presentation plus multiple-choice questions)?
3. teacher modification of the on-going presentation is guided by feedback of subjects' answers to relevant interpolated multiple-choice questions (regular presentation plus multiple-choice questions plus program analyzer--PAR--feedback)?

Nine teachers each gave 20-minute live televised lessons over three content areas--phonetics, propaganda, and parliamentary procedure. Each was given the appropriate set of behavioral objectives for the content he was to teach, as well as a corresponding set of relevant ("typical") multiple-choice questions which might be used to measure these objectives. (These questions were selected from the pool of retention test items described previously.) Each individual instructor was asked to prepare his own lecture around the objectives and questions, but was permitted to arrange the sequence and present the topic in any fashion he considered appropriate within the 20-minute limit. The variation in method of instruction from this point on consisted of the following.

1. Regular presentation. The multiple-choice questions were not presented to the subjects during any part of the lecture, but were converted into statements by using stem and answer together. These statements were interpolated at appropriate locations chosen as indicated below.
2. Regular presentation plus multiple-choice questions. The multiple-choice questions were asked by the instructor shortly after the related points were made in the lecture; the only other restriction was that they be periodically distributed. No report of student answers was made to the instructor.
3. Regular presentation plus multiple-choice questions plus program analyzer (PAR) feedback. The multiple-choice questions were again periodically asked after the related lecture points; in this case, however, reports of subject answers were made to the instructor within 10 seconds after he had asked the questions. This information was collected by the PAR and presented in terms of numbers of subjects selecting each of the five choices. Based on this, the instructor was in a position to modify his presentation.

Each instructor was randomly assigned to a randomly replicated Graeco-Latin square design condition in order to equate instructor differences and order effects in the experiment. The basic design of each of the three replications is presented in Figure 1. A treatments (methods) X treatments (contents) X replications (random groups) analysis was used.

Contents**

		Ph	Pr	Pa
Methods*	Reg.	a,2***	c,3	b,1
	MC	b,3	a,1	c,2
	PAR	a,1	b,2	a,3

* Methods:

Reg. - regular televised presentation
MC - regular and multiple-choice questions
PAR - regular and multiple-choice questions and program analyzer feedback

** Contents:

Ph - Phonetics
Pr - Propaganda
Pa - Parliamentary Procedure

*** a,b,c, represent 3 instructors in a replication; 1,2,3 represent the order of presentation.

Figure 1. Basic Graeco-Latin Square Design Used in Experiment I. A. Speech.

Twenty-seven groups (classes) of 13 to 15 freshmen speech students, serving as subjects, were randomly assigned viewing conditions (subjects were dropped randomly to leave 13 in each cell). All viewed the lectures in the same room over the same television monitor. Following the presentation, the 30-item immediate retention test described above was administered.

Table 4 shows the results of the experiment. As noted, only the Content variable was significant. Since this may be due to differences in the difficulty level of the tests used with the three contents, this finding is of little importance. The terms of greater interest, the Methods and the Methods by Content interaction terms, yielded no significant F's.

Table 4

Analysis of Variance of Speech Experiment I. A.

Source	df	SS	ms	F
Methods	2	84.52	42.26	<1
Content	2	6,909.85	3,454.93	9.50*
Replications	2	525.85	262.93	
Content x Methods	4	634.815	158.70	<1
Rep. x Content	4	1,434.81	363.70	
Rep. x Methods	4	496.82	124.20	
Rep. x Cont. x Methods	8	4,023.19	502.90	

* $p < .05$

Since the analysis may be rearranged to make the Order (first, second, or third television presentation) a systematic variable rather than a control variable that was treated as a random variable, a post hoc analysis was conducted to determine whether the effects of this variable were significant. An analysis similar to the above, but with order and content as the two main variables, was conducted. Neither of the terms involving order was significant (order, F = 1.11; order x content, F = 1.84). The mean numbers of items correct for the three main effects investigated are shown in Table 5.

Table 5

Mean Numbers of Items Correct in Speech Experiments I. A. and I. B.

Television Presentation Results						Audio Tape Results	
Method		Contents		Order		Method	
Reg.	18.68	Phonetics	17.89	1st	18.12	Reg.	19.18
MC	18.42	Propaganda	20.34	2nd	18.55	MC	18.18
PAR	18.72	Parlia.Pro.	17.60	3rd	19.16	PAR	17.92

Fall, 1962: Experiment I. B. Speech. In addition to the televised lessons above, audio tapes of nine lessons were made and used on nine additional sections of Speech 1 students. These tapes were produced in the same way as the televised lessons. Three graduate students who had not participated in the televised lessons were randomly assigned to contents. Each taught one lesson under each method, in a random order. Teacher and content were thus confounded. The same three methods variables were used, and the three contents-teacher combinations were treated as random replications, making a 3 x 3 treatment by replications design. Using audio tape, the methods effect was still non-significant. The method means are shown in Table 5, above, and the analysis in Table 6.

Table 6

Analysis of Variance of Speech Experiment I. B.

Source	df	SS	ms	F
Methods	2	34.38	17.19	<1
Replications	2	61.04	30.52	
Rep. x Methods	4	128.29	32.38	

Fall, 1962: Experiment I. C. Statistics. This experiment was designed to answer the same three basic questions asked in regard to the speech area above, but the experimental plan was different.

Four instructors each presented two 50-minute television lectures. Four content areas, each divided into two parts, were used: frequency distribution (parts I and II), central tendency (I and II), variability (I and II), and normal curve (I and II). Each instructor was given the appropriate set of behavioral objectives for his content area, plus a corresponding set of relevant ("typical") multiple-choice questions which might be used to measure these objectives (the questions being selected from the item pool described earlier). As before, each instructor was asked to prepare his own lecture around the objectives, but was allowed to arrange the sequence and present the topic in any fashion he considered appropriate within the 50-minute limit. The "typical" items for each presentation were given to the instructor only after a rehearsal of that particular lesson, and he was instructed that his presentation should not be oriented toward these particular items. The same variations in method were used here as in the speech experiment: regular presentation; regular presentation plus multiple-choice questions; regular presentation plus multiple-choice questions plus program analyzer (PAR) feedback. However, the operational implementation of these methods was slightly different. When the acting instructor reached the point in his presentation where he wished to have the typical questions presented to the students, he pressed a button on the lectern. The regular course instructor was seated at a monitor with a microphone and with another television camera nearby. When the button was pressed, the studio engineer switched to the other camera which showed a card containing the "typical question" in its statement form or as a multiple choice item. If the former, the regular instructor simply read the item from the card. If the latter, he read the item from the card, waited 15 seconds, and then read the correct answer. If the PAR condition was being used, he pressed the PAR button to start counting responses after 12 seconds had elapsed. The acting instructor could then observe the responses made while the correct answer was being read.

Each instructor, with each of his contents, presented one-third of his material by means of the regular method, one-third using the regular plus multiple-choice, and one-third using the regular plus multiple-choice plus PAR feedback presentation method. The basic treatments X replications design planned is presented in Figure 2.

Replications**	Methods*		
	Reg.	MC	PAR
FD I			
FD II			
CT I			
CT II			
VAR I			
VAR II			
NC I	XXXX		
NC II	XXXX		

* Methods

Reg.= regular televised presentation

MC = regular + multiple-choice questions

PAR = regular + multiple-choice questions + program analyzer feedback

** Replications:

FD = frequency distribution

CT = central tendency

VAR = variability

NC = normal curve

Figure 2. Basic Treatments X Replications Design Planned for Experiment I. C. Statistics.

The same class of approximately 40 students served as subjects for each replication and observed all three methods of presentation for each. A random 15 students viewed the lectures in the PAR room adjacent to the television studio and with one monitor, while the remainder watched from two similar monitors in the normal classroom. Following the presentation, the students had a 5-minute break, and then a 50-minute test on the material was administered. The students were told these tests would be summed at the end of the semester, and the final summed grade on these would count as one of the regular course tests. They seemed to take these tests seriously and to work hard at them.

Unfortunately, due to failures in the equipment, in one case, and a failure of the instructor to meet the specified time limits in another case, the two cells containing X's in Figure 2 were lost. The analysis thus had to be broken down into two parts as shown below.

Results: I. C. Immediate Post Tests. Since all subjects saw all presentations, but different test items were needed to test on each content, the unit of sampling in this design was the test item. The proportion of correct responses for each test item was determined and averaged to yield the mean proportion of correct responses for each method on each topic (the mean proportion of correct responses for all items corresponding to each cell of Figure 2). Using this as the basic response measure, a treatment by replications design was conducted over the three treatments for the first six replications. A similar comparison was made for the PAR and MC methods over the entire eight replications. The analyses are shown in Table 7.

Table 7

Analyses of Variance of Statistics Experiment I. C. Post Tests

3 x 6 Analysis				2 x 8 Analysis			
Source	df	ms	F	Source	df	ms	F
Methods	2	.00727	<1	Methods	1	.00369	<1
Rep.	5	.06763		Rep.	7	.03448	
R x M	10	.00967		R x M	7	.01139	

Since the order of methods was randomly determined for each replication, any order effects were assigned as a random source of variance to the error term (R x M). It occurred to the experimenters later that, in any given presentation of a sequential subject matter like statistics, the first material tends to be easier than the last material presented. An analysis was therefore run as a 3 x 6 treatments by replications design to test this hypothesis. The order effect was not significant (see Table 8), although the mean proportion correct did decrease from the first to third parts of the presentations.

Table 8

Treatments x Replications Analysis of Variance

of Statistics Experiment I. C. Post Tests

Source	df	ms	F
Order	2	.0061	<1
Rep.	5	.0676	
O x R		.0099	

Results: I. C. Retention Tests. In addition to the immediate post tests, items on the behavioral objectives covered by the television presentations were inserted into the regular course examinations. None of the items so used was identical to any used in the television presentation as "typical items" or any used in the immediate post test. A total of 102 items were inserted in the course tests to test the retention of behavioral objectives covered in the presentations. In order to include the two lecture periods that had been deleted from the two normal curve presentations due to equipment failure and instructor error, the course instructor (Games) lectured on the missed objectives for the scheduled amount of time in the next class period. An additional dimension was added to this analysis in that the printed behavioral objectives that had been given to the instructor were also distributed to the class for one of each of the two presentations. A coin toss decided which set of objectives was given, and the set of objectives covered in one of each pair of presentations was given to the class after both presentations were complete, but prior to the regular course test. The analysis thus had three dimensions: The three methods of presentation, printed behavioral objectives received vs. no behavioral objectives received, and the four content replications for a 3 x 2 x 4 analysis. Again the mean proportion of correct responses per item was used as the basic behavioral measure per cell. Since 102 items were used, and 24 cells, the mean number of questions per cell was only 4.25. Unfortunately, due to uneven numbers of items available per objective and time pressures that made it impossible to write more as needed, some cells were based on the mean proportion correct of only two items; the modal number of items per cell, however, was 5. In the analysis, it was found that the three interaction terms involving replications (the 3 error terms) were homogeneous ($F_{max} = 1.36$) so they were combined into a single error term, yielding the analysis presented in Table 9. As can be seen, the analysis of retention data failed to show any significant effects.

Table 9

Analysis of Retention Tests Using Mean Proportion of Correct Responses, Experiment I. C. Statistics

Source	df	ms	F
Objectives	1	.0007	<1
Methods	2	.0058	<1
Replications	3	.0091	
Objectives x Methods	2	.0080	<1
Error	15	.0112	

Fall, 1962: Experiment I. D. Statistics. One other minor experiment was conducted in statistics classes in the fall of 1962. Both an undergraduate and a graduate class taught by the regular course instructor (Games) study correlation and regression at the end of the semester. Behavioral objectives in correlation and corresponding test items were available, so the behavioral objectives were divided into short blocks on the basis of organization, each block being essentially a short unit covered on the topic. These blocks were randomly divided into two halves, A and B. The graduate class was given dittoed sheets for the A objectives, and the undergraduate class was given dittoed sheets containing the B objectives. The final examination of both classes contained the same items over the entire set of objectives, 15 items on behavioral objectives A and 15 items on behavioral objectives B. Considering each item as a random sample from the population of items for that objective, it was again possible to use the item as the unit of sampling and conduct a Type II (Lindquist, 1953, p. 273)[15] analysis of variance on the 2 x 2 design presented in Figure 3 below. The results are indicated in Table 10. The only significant effect was class, yielding the expected finding that the graduate students (as a group) scored higher than the undergraduate students.

		OBJECTIVES	
		Received	Not Received
CLASS	Graduate	A	B
	Under-graduate	B	A

Figure 3. Design Used in Experiment I. D. Statistics.

Table 10

Analysis of Variance of Statistics Experiment I. D.

Source	df	SS	ms	F
Between items				
Objectives x class	1	.0070	.0070	<1
Error (b)	28	1.6949	.0605	
Within items				
Objectives	1	.0020	.0020	<1
Class	1	.1242	.1242	13.575*
Error (w)	28	.2562	.0092	

* p < .05

---

[15] Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. New York: Houghton and Mifflin Co., 1953.

Fall, 1962: Summary. In this section, various conclusions based on the first semester's work are presented. Emphasis has been placed upon those considerations which led to revisions in the original second semester's research plans; other more general considerations have largely been reserved for the major Discussion section to follow later in the report.

One of the methods used in an attempt to evaluate possible methods of presentation in the beginning statistics course was a questionnaire given at the end of the course (see Appendix C). Students were asked to choose the techniques they would prefer during class hours for the following two course purposes: 1. communicating course objectives; 2. teaching course objectives (desired behaviors). The total number of "mentions" in Table 11 is greater than the number of student respondents, which was 29; the total number of "first choices" is slightly smaller because several students failed to provide a first choice.

Table 11

Results of Fall, 1962, Questionnaire on Preferred

Methods of Presentation in Statistics Course

	Course Purposes			
	Communicating Objectives		Teaching Objectives	
Method of Presentation	Mentions	First Choices	Mentions	First Choices
1. Small lecture (40 students); questions possible and encouraged	25	22	28	21
2. Television or film presentation	10	4	10	2
3. Large lecture (250 students); no questions possible	8	2	1	1
4. Small-group discussion with graduate student	5	0	12	3
5. Class discussion (40 students); instructor answers questions only (no additional lecturing)	4	0	5	0

In addition to the data presented in Table 11, other data showed some support (preference) for the use of programs (teaching machine type) during class hours for the teaching of objectives. Class discussion, as a method, received less support than each of the following programing methods: small-step programs (6 mentions, 1 first); large-step programs (8 mentions, 0 firsts); doing exercises in class (6 mentions, 1 first). It is difficult to evaluate this type of information, particularly in the absence of a control group. However, several observations are worth making:

1. The status quo method--small lectures--was clearly preferred by this class.
2. The television or film presentation method was mentioned as a desirable method by approximately one-third of the respondents.

The above study also showed some student preference for a television or film method over a large lecture method, particularly for the teaching of objectives (course behaviors). This suggested that the kind of research-based television presentation used here deserved further exploration. However, three major changes were made in subsequent work.

First, film was substituted for the television presentation. The practical difficulties of repeating the same "method" differences on different contents throughout the same course (as in statistics) were considerable. Therefore, a permanently-recorded filmed presentation was made for the statistics contents. It could be modified as necessary for the various method changes desired, yet could be relied upon as presenting the same procedures to the large numbers of subjects to be used in subsequent statistics work.

Second, the investigation of ways in which the teacher might use "typical items" was dropped. The absence of any significant effects due to the different methods used suggested that this line of approach was likely to continue to be unprofitable. Analysis suggested that the teacher and the typical items often seemed to serve as an intermediary for translating the behavioral objectives to the student. That is, the behavioral objectives were given to the teacher, and both he and the typical items translated them into examples for the student. The problems raised by most attempts at translation in other situations suggested that this indirect communication might be a hindrance to learning. It seemed that giving the student direct exposure to the objectives, in addition to the teacher's efforts, might be preferable. This method was adopted in subsequent work.

Third, the use of the behavioral objectives during breaks in the presentation was adopted. Previous work had shown breaks themselves to be desirable (see Discussion section), and the behavioral objectives seemed likely to be the best material to interpolate. It was hoped that this combination would improve learning both by providing a change of pace in the presentation and by providing "organizers" to which material could be tied.

Spring, 1963: Experiment II. Speech. This experiment was designed to answer the following questions:

Will subjects learn more from a recorded (audio) presentation when behavioral objectives are:

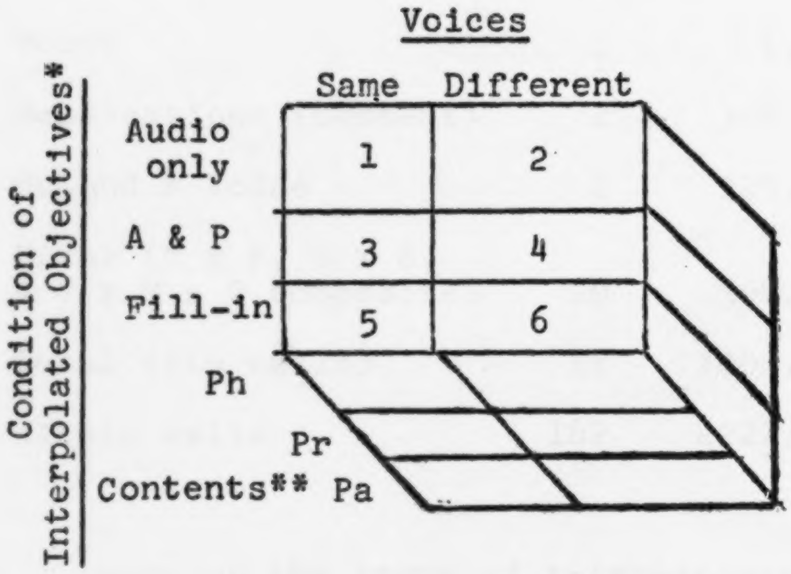
1. a) presented orally only, both orally and in printed form, or orally with certain words omitted?
   b) presented by means of the lecturer's voice or a different voice?

2. a) presented or not presented orally?
   b) presented or not presented in printed form?

One teacher gave 20-minute lectures over each of three content areas--phonetics, propaganda, and parliamentary procedure. As before, he was given a set of behavioral objectives for the content he was to teach, and he was again permitted to arrange the sequence and present the topic in his own fashion. In this experiment, however, the lecture was recorded on audio tape and presented to the subjects through the audio portion of the television set. The tapes were made in the following manner for each content.

Two tapes were recorded simultaneously; on both, the objectives sections were retained, but on the one certain words were eliminated by cutting the volume below recording level. Three additional tapes were then made from these in which the entire objectives section was eliminated by reducing the volume. For one of these, objectives read by a different voice were dubbed in; for another, objectives read by the different voice were dubbed in, but the same words as above were eliminated; for the third, no objectives were dubbed in (a "dead air" condition). These tapes were then used alone or in conjunction with printed objectives that were read during the "objective" times. The revised 30-item retention tests were given immediately after the tapes were completed.

The basic design of this study is presented in Figure 4, and the analysis of results in Table 12. Twenty-four sections of approximately 10 to 14 Speech 1 students were available. These sections were randomly assigned to 24 conditions composed of the 18 found in this analysis of the experiment, plus six additional conditions shown in the next analysis of the experiment. Subjects were randomly dropped from the Speech 1 sections to make ten subjects in each cell; however, this was done merely to check that the variance within cells was smaller than the variance associated with the proper error terms. The section was actually the random sampling unit, since any variation in conditions of timing of the test or of temperature, etc., would act on the entire section.


Voices
Same
Different
Condition of Interpolated Objectives*
Audio only
A & P
Fill-in
1
2
3
4
5
6
Ph
Pr
Contents** Pa

*Condition of interpolated objectives (Methods):

Audio only - oral presentation only

A & P - oral plus printed presentation

Fill-in - oral presentation with words missing-students requested to write the missing words on an answer sheet.

**Contents:

Ph - Phonetics

Pr - Propaganda

Pa - Parliamentary procedure

Figure 4. First Design Used in Vary Presentation of Objectives in Experiment II. Speech.

Results: Experiment II, First Analysis. The results of this analysis of the test scores are shown in Table 12. Since the interaction terms involving replications (contents) were homogeneous, they were combined into a composite error term to increase the degrees of freedom.

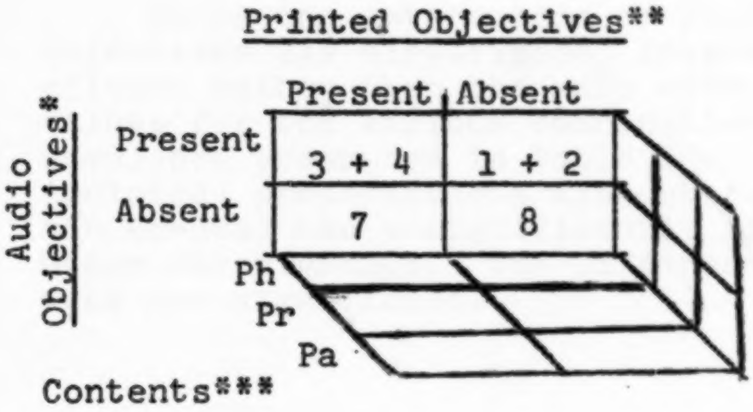
Table 12

Analysis of Variance of Experiment II. Speech; First Analysis

Source	df	SS	ms	F
Methods	2	114.63	57.31	1.436
Voice	1	1.09	1.09	<1
Replications (Content)	2	366.23	183.12	
Method x Voice	2	125.68	62.84	1.575
Error (M x R, V x R, & V x M x R composite)	10	398.97	39.90	
Total (btn cells)	17	1006.6		
Within cells	162	2222.2	13.717	

None of the terms of interest was significant when using the proper error term. The fact that the Voice term and the Methods x Voice interaction would be significant when using mean square within cells as an error term is an example of the necessity of being sure that the error term reflects the types of random variation actually occurring in the experiment.

Results: Experiment II, Second Analysis. For the second analysis, the data present in the upper four cells of the previous analysis were combined with the data from the other six presentations to make a treatments (2) by treatments (2) by Replications (contents-3) analysis. This analysis covered the effects of using combinations of audio and printed presentations of the behavioral objectives during the "breaks" of the audio tape presentation. Since the Voice condition had had a negligible effect (means of 13.72 and 13.88 on the revised test), the two Voice conditions (same and different voices) were collapsed into a single cell of 20 subjects. The design for this analysis is shown in Figure 5.



* - oral presentation of objectives used or not used

** - printed presentation of objectives used or not used

*** - Contents

Ph - Phonetics
Pr - Propaganda
Pa - Parliamentary Procedure

Figure 5. Second Design Used in Varying Presentation of Objectives in Experiment II. Speech.

In this situation, significant results were obtained, as shown in Table 13. Both the main effect of the use of printed objectives and the interaction term of printed vs. oral presentation of objectives were significant. The mean square within cells, which was partially inflated by combining two sections in the Audio-Present group, was larger than the other error terms. In Table 13 below, this term is used as a means of providing a conservative test (all F's would be larger using the usual error terms.)

Table 13

Analysis of Variance of Experiment II. Speech: Second Analysis

Source	df	SS	ms	F
Audio Objectives	1	21.51	21.51	1.381
Printed Objectives	1	137.81	137.81	8.846*
Printed obj. x Audio obj.	1	96.10	96.10	6.169*
Replications (contents)	2	289.94	144.97	
Combined error term (Pooled RxP, RxA, RxAxP)	(6)	(29.15)	(4.86)	
Within cells	168	2,617.26	15.58	

*$p < .05$

Since the interaction of Paper and Audio presentation of objectives was significant, interest centered on the simple effects rather than the main effects. The mean test score values for the various combinations, based on the 30-item test used, are presented in Table 14. The group having Paper (printed) presentations without Audio presentation (with dead air breaks) had a significantly higher mean than any of the other three groups. The differences between the other groups were not significant.

Table 14

Mean Test Scores for Groups Having Different Presentations of Behavioral Objectives in Experiment II. Speech: Second Analysis

Method of Presentation	Test Score
1. No Paper Objectives and No Audio Objectives	13.2
2. Paper Objectives but No Audio Objectives	17.0
3. No Paper Objectives but Audio Objectives	14.0
4. Paper Objectives and Audio Objectives	14.7

Statistics, Spring, 1963

Behavioral objectives had previously been written for several sections, with corresponding test items for each objective. To obtain item analysis information for improvement of the tests in the coming fall experiment, and to test the effect of the usage of the behavioral objectives, the behavioral objectives on Frequency Distribution, Central Tendency, and Normal Curve topics were prepared for testing in two sets; one set had the odd-numbered objectives, and the other had the even-numbered objectives. The 9 A. M. class of Psychology 105, Elementary Statistics, received one set, and the 2 P. M. class received the other. The closed-book portions of regular course tests contained test items on the entire set of objectives. The instructor used the best two items for each objective (in his judgment) and randomly placed one item in the 9 A. M. test and the other item in the 2 P. M. test. Thus, the test items were randomly assigned to the "Objectives received" or "No objectives received" condition. Again, since each subject took all items in a given test, the mean proportion of correct responses under a given condition is a linear transformation of the sum of correct responses across subjects; therefore, the analysis was carried out as a treatment by replications design on mean proportions of correct responses. The two classes used as described above yielded six replications. In addition, the correlation and regression material was also split into odd and even sections, and one half of the objectives were given to all subjects. This section was covered on a common final, which was taken by both classes at the same time. The correlation items were treated as one replication, and the regression items were treated as another replication, making a total of eight replications. The analysis of the data follows:

Table 15

Analysis of Variance of Proportion of Correct Responses on Eight Replications

Source	df	SS	ms	F
Objectives received	1	.0032	.0032	<1
Replications	7	.0247	.0035	
Replications x Objectives	7	.0304	.0044	

Although the difference was in the expected direction (mean proportion of correct responses was .713 for the objectives received, and .684 for those not received), the difference was small and was clearly non-significant.

Summer, 1963: Experiment III. Statistics. A major objective of the summer was to improve the behavioral objectives on the statistics topics of Frequency Distributions, Central Tendency, and Variability; a related objective was to construct instructional films on these objectives, and achievement tests on the topics covered by the films. All items on these topics that had been used the previous year were item-analyzed, and many new items were written. The subjects available were limited and were used for item tryout and experimentation.

The first investigation used an intact class of General Psychology students (on July 15) after the students had been assigned the statistics chapter in their text and had had the frequency distribution portion covered in lectures, but before they had received a formal test on this material. All materials were randomly ordered and placed on the desks in the classroom. Students sat at their usual seats.

The materials each student received consisted of a 46-item pretest, a set of behavioral objectives, and a 53-item multiple choice posttest (with a corresponding IBM card answer sheet, and a sponge and stylus for self-punching). The instructions used on the pretest and posttest are included in Appendix C.

The pretests were prepared in four different versions. An analysis suggested that the behavioral objectives for the Frequency Distribution film might have a hierarchical order such as that found by Gagne and Gagne and others.[16] The 17 objectives were broken down into seven levels, with level one judged to contain the simplest tasks, level two somewhat more complex tasks, possibly requiring knowledge of those in level one for successful completion, etc. Level seven was the most complex. Success at many of the tasks in lower levels was judged to be necessary in order to successfully carry out the tasks called for in level seven. For each objective in each level, at least one open end (OE) and at least one multiple choice (MC) item was included in the pretest. It was hypothesized that if students took the pretest proceeding from level one to level seven, they would be following an ascending level of complexity desirable for teaching and would "learn more" from the test than if they followed the opposite sequence from seven to one. Similarly, since the MC questions normally involve merely recognizing the correct answer and have a higher probability of eliciting correct answers than OE questions, it was felt that taking MC items first would be more beneficial to learning than taking OE items first. Consequently, four different forms of the pretest were used, in the following combinations.

---

[16] Gagne, Robert M. "The Acquisition of Knowledge." _Psychological Review_, 1962, _69_, 355-365; Gagne, Robert M., Mager, John R., Aarstens, Helen L., and Paradise, Noel E. "Factors in Acquiring Knowledge of a Mathematical Task." _Psychological Monographs_, 1962, _76_ (Whole No. 526), 20 pp.

		Hierarchy of Pretest Items	
		7 → 1	1 → 7
Order of Pretest Items	MC first		
	OE first		

Figure 6. Combinations of Pretest Items Used in Experiment III. Statistics.

Thus, the pretest represented in the upper left-hand corner consisted of a multiple choice section starting with level seven first, then levels 6, 5, 4, 3, 2, 1, then an OE section, again going from seven to one. Once the student turned a page, he was not to turn back to it again.

In addition to the four conditions created by the use of different pretests, three different sets of behavioral objectives were given: the regular frequency distribution objectives, presented in what was judged to be a desirable sequence, generally running from simpler levels to higher, a set of scrambled objectives formed by randomizing the order of the complete sentences in the objectives, and a set of variability objectives, which consisted of extraneous material as far as the Frequency Distribution test was concerned. It was originally planned to divide the 88 students in the class into about 44 with the regular objectives and 20 with each of the other sets of objectives. Due to the randomization procedure used, and to several absences from the class, these goals were only roughly approximated. The following adjustments were necessary: (a) Two students were dropped for suspicion of cheating and their papers were not scored; (b) In two cells of the three-dimensional table, one subject was chosen at random and dropped to get proportional n's; and (c) In two cells, the mean was added as an extra case, again to get proportional n's (Lindquist).[17]

The subjects were given only 30 minutes for the pretest. These data were not analyzed since this time limit proved short and most subjects had not finished all of the items. Since the MC items have a higher chance expectation than the OE, and since the level one items tended to be less difficult than the level seven, it was felt this analysis could only yield spurious results.

At the end of 30 minutes the pretest was collected, and the students had 10 minutes to study the sheets of objectives.

---

[17] Lindquist, op. cit., p. 148.

These were then placed under the seats, face down, and the experimenter read the instructions on the final test and on the use of the IBM answer card (with the students). This took two minutes, so that students then had 32 minutes to complete the multiple-choice posttest. A few left early, but most took the entire time and several failed to finish the last few items. Several of the items on the test permitted multiple responses (see Appendix C). A scoring key was constructed so that the best "score" possible was -10 and the poorest "score" possible was 86. This measure is referred to as "number of errors" for convenience of communication. The analysis of this measure is shown in Table 16.

Table 16

Analysis of Variance of Number of Posttest Errors, Experiment III. Statistics

Source	df	SS	ms	F
Objectives	2	1786.42	893.21	9.88**
Order of pretest	1	2.09	2.09	<1
Hierarchy of pretest	1	91.19	91.19	1.01
Obj. x Hier.	2	146.28	73.14	<1
Obj. x Order	2	114.75	57.38	<1
Order x Hier.	1	1.13	1.13	<1
Obj. x Order x Hier.	2	114.10	57.05	<1
Within Cells	62	5604.37	90.39	

**p < .01

The significant Objectives main effect was clearly due to more errors for the subjects who had studied the inappropriate Variability objectives, as opposed to those who had studied the appropriate Frequency Distribution objectives. The mean number of errors for the three objectives conditions were 22.57 for the normal hierarchical behavioral objectives, 22.61 for the scrambled sentences from the behavioral objectives, and 33.58 for the inappropriate Variability objectives. Apparently the opportunity to study pertinent objectives for 10 minutes reduced the number of errors by about 11; however, whether the objectives were presented in what the investigators judged to be a desirable hierarchical order as opposed to a random order had a negligible effect ($t < 1.0$). This result is similar to that obtained by several investigators who have found that subjects do as well on scrambled learning programs as they do on original sequential programs.[18]

A smaller experiment was conducted with the Central Tendency items using volunteer subjects from an Educational Psychology class. However, the number of volunteer subjects was so small that the power of the test was negligible. The data were, however, used for item selection.

Fall, 1963: Experiment IV, Speech. The previous experiment and other programing experiments using visual stimuli suggested that much of the type of content learning in educational experiments may be similar to paired-associates learning. In this case, the pairs may be words, or phrases, or combinations of these, rather than nonsense syllables, but the essential responses measured may still require the forming of paired associates. The speech experiment was designed to explore related aspects of this notion with respect to the speech content objectives and tests as used in the spring of 1963. Unfortunately, the design was not carried out exactly as planned, but the following questions were asked.

Will subjects learn more from a recorded (audio) presentation when behavioral objectives are:

1. presented orally, or not presented?
2. presented in printed form, or not presented?
3. presented with "key" words in capitals (printed form), or presented as the key words only (oral or printed form)?

---

[18] Johnson, Nancy J., "A Study of Linear Programmed Speech Content Using Rigid and Random Sequences." Unpublished master's thesis, Ohio University, 1963; Jacobs, Paul I. "Some Relationships Between Testing and Auto-Instructional Programing." Audiovisual Communication Review, 1962, 10, 319; Roe, K. V., Case, H. W., and Roe, A. "Scrambled Versus Ordered Sequence in Autoinstructional Programs." Journal of Educational Psychology, 1962, 53, 101-104.

One teacher gave 20-minute lectures over each of the content areas--phonetics, propaganda, or parliamentary procedure. Behavioral objectives to cover the area were provided, and the instructor arranged his own sequence and presented his topic in his own fashion. As in Experiment II. Speech, audio tapes were used to present the lesson to the subjects. The tapes were made in the following manner for each content.

Two tapes were recorded simultaneously, with the objectives section intact on both. Two additional tapes were then made from these. In one, the objectives section was left blank by reducing volume (dead air condition), while in the other "key" words were read. These words were chosen by combining the independent choices of four judges (staff members) who attempted to pick the key words or word phrases in the objectives, i.e., those words which carried the "meaning" of the objectives. From two to six key words (average, approximately four) were selected from objectives which averaged about 20 words in length in regular (sentence) form. Sixty per cent of the objectives statements for each content were handled in this manner; the other forty per cent were not readily amenable to this kind of analysis and were left in sentence form. These tapes were then used in conjunction with printed objectives that were read during the "objectives" time. In one printed form, the objectives were presented in regular (sentence) arrangement; in a second, the key words were capitalized, and in the third, the key words were isolated and printed in capitals. (As indicated previously, these objectives appear along with other versions in Appendix A.) Figure 7 presents the various treatment combinations.

		Condition of Print Objectives			
		No Print obj.	Capitalized Key words in objectives	Regular objectives	Key Words only
Audio Conditions	Audio reg. obj.	1 13.75	2 15.19		
	Dead Air	3 12.89	4 16.22	5 13.97	6 15.56
	Audio Key Words only			7 17.00	8 15.50

Figure 7. Treatment Conditions Used in Fall, 1963, Speech Experiment and Mean Number of Correct Responses for Each Treatment Averaged Over the Three Contents.

Class groups of 12 to 15 subjects from Speech 1 again served as the unit of sampling, with the previous three contents serving as three replications. Each class was given a retention test immediately after the presentation. The tests were the same 30-item tests used in the spring, 1963, speech experiments. Subjects were randomly eliminated from classes with more than 12 subjects, leaving 12 measures in each cell.

Analysis of Results

The cells numbered 1 to 4 in the upper left hand corner were first analyzed in a 2 x 2 x 3 replications analysis. The means of each cell are shown in Figure 7. The analysis is in Table 17. The replication interaction terms were homogeneous ($F_{max} = 1.78$) and were combined into a single composite error term.

Table 17

Analysis of Variance of Treatment Conditions Used in Fall, 1963, Speech Experiment

Source	df	SS	ms	F
(A) Reg. Audio Obj. vs. Dead Air	1	.25	.25	<1
(P) No Print Obj. vs. Capitalized Obj.	1	205.44	205.44	8.21*
Interaction	1	32.11	32.11	1.28
Replications (contents)	2	32.76	16.38	
Composite Error term (Pooled A x R, P x R, + A x P x R)	8	200.24	25.03	
Within cells	132	2817.16	21.34	

* $p < .05$

As in the previous spring(Speech Experiment II, second analysis), the main effect of Print objectives vs. No Print objectives was significant; however, although the interaction was in the same direction, in this analysis it was not significant. The great similarity of the two results is shown by graphing the results of the two experiments in Figure 8. It should be noted that the printed objectives in this experimental comparison had the key words capitalized, while those in the spring, 1963, experiment had the regular objectives. In both experiments it is clear that the use of the printed objectives substantially increased test performance. The mean score in this analysis for the "No printed objectives" groups was 13.32, while those with objectives had a mean score of 15.71. Although this difference of 2.39 may not be considered as very large, it must be remembered that the test used consisted of only thirty items and lasted only twenty minutes. Similarly, the duration of the lesson in which the treatment differences were imposed was only twenty minutes. An increase of 18% (using the lower group as a base level) in correct responses would seem to be rather substantial. (The increase was 17% for the same comparison in the spring, 1963, experiment.)

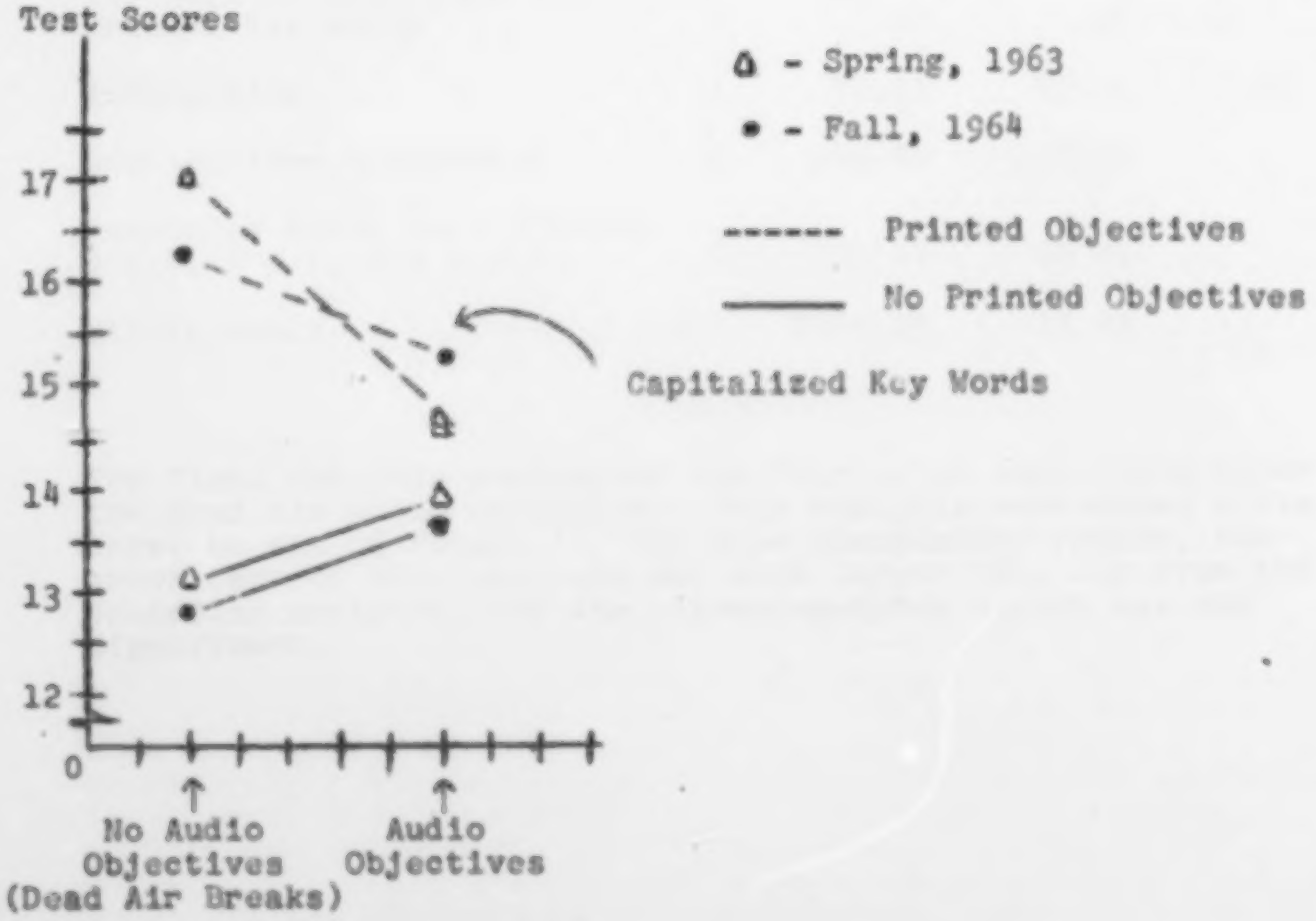


Figure 8. Comparison of Results of Spring, 1963, and Fall, 1964, Speech Experiments.

A similar analysis was conducted for the four cells in the lower right hand corner, numbered five to eight. None of the terms of interest was significant. This analysis is found in Table 18. The hypothesis of homogeneity of the replication interaction terms was retained ($F_{max} = 4.64$) so a pooled error term was again used.

Table 18

Analysis of Variance of Certain Data in the

Spring, 1963, Speech Experiments

Source	df	SS	ms	F
(D) Dead Air vs. Audio Key Words	1	76.56	76.56	1.53
(K) Reg. Printed Obj. vs. Printed Key Words	1	.17	.17	<1
Interaction	1	82.51	82.51	1.65
Replications (contents)	2	249.06	124.53	
Composite Error term (Pooled R x D, R x K, R x D x K)	6	229.44	49.91	
Within cells	132	2960.25	22.43	

The final analysis contrasted the four print conditions under the dead air audio condition. This analysis contrasted cells three to six of Figure 7. For some unexplained reason, the error term of this analysis was much larger than any from the preceding analyses, and the between-methods effect was not significant.

Table 19

Analysis of Variance of Certain Data in the

Spring, 1963, Speech Experiments

Source	df	SS	ms	F
Print conditions	3	246.69	82.23	1.01
Replications (contents)	2	33.10	16.55	
R x Print	6	489.46	81.58	
Within cells	132	2691.08	20.39	

Experiment IV. Statistics. This experiment was designed to answer the following questions.

Will subjects learn more from a filmed presentation in which an initial study period is used and:

1. which contains a pretest, or behavioral objectives, or extraneous materials?
2. where the following (filmed) lesson contains interpolated objectives, or dead air, or no breaks in time?
3. retain longer when the learning period is or is not followed by different learning?

Basic presentations were made of two content areas--frequency distribution and central tendency (as previously noted, a video tape was made of a live presentation, with the tape converted to film). The films were then prepared in two versions, one with silent time breaks and another with no time breaks. The time breaks were inserted at points in the film immediately after objectives had been covered, and were timed so that the subjects would be able to read the appropriate objectives during the break. This copy of the film was then used in two conditions, one arranged so that the breaks simply left a period of no sound and no light on the screen (dead air), and the other with a slide projector coordinated so that the appropriate objective was shown on the screen during the break. The film was cued so an assistant at the projector saw the cue and pressed a button to activate the slide projector, while pre-set timers controlled the automatic projector so the visual presentation of the slide coincided with the time break. These variations then constituted what we shall call the Presentation variable:

P1) film with visual objectives interpolated in breaks;
P2) film with "dead air" breaks; and
P3) continuous film, with no breaks.

Prior to the filmed presentation, all subjects had a 12-minute period in which they engaged in one of three activities.

1) Took a pretest on the material to be covered. Items were arranged in a hierarchical sequence from the least complex to most complex.

2) Studied the list of behavioral objectives for the topic to be covered.

3) Studied extraneous material, of a high degree of "face pertinence," but containing material not essential to the task ahead. The material was a dittoed section on "Mathematics prerequisite to statistics."

These conditions will be referred to as the Introduction (I) variable. These two dimensions (P and I) were used in experimental sessions on both Frequency Distributions (F. D.) and Central Tendency (C. T.). Subjects run were students from Introductory Psychology classes who had not studied the statistics chapter of the introductory text or covered any of this material in class prior to completion of the experiment. The subjects signed up for two-hour sessions at the same period for three successive weeks. They were run according to one of the three following schedules.

	Week 1	Week 2	Week 3
Group A1	F.D.film and test	C.T.film and test	Retest on F.D.
Group A2	F.D.film and test	----------	Retest on F.D.
Group B	C.T.film and test	F.D.film and test	

The retention tests of week three were conducted by sending the students to a testing room under the direction of departmental graduate assistants. There they were given one hour to complete the same F. D. test that was completed previously. All other sessions were run in a small classroom set up for this experiment. Subjects were signed up in small groups, from 3 to 20 subjects per group. Groups were randomly assigned to Presentation (P) conditions, and subjects were randomly assigned to Introduction (I) conditions within each group. During the second week, subjects in the A1 and B groups saw the other film under the same condition as the week before. The schedules for group sessions were 12 minutes on introductory materials, 40 minutes watching the film, 5 minutes rest, 55 minutes maximum on the posttest. The posttests were in two parts, an open-end test that was completed first, then a multiple-choice test that was completed after the open-end test was handed in. In general, both open-end and multiple choice tests were arranged in a hierarchical sequence, from the early objectives (least complex) to the last objectives (most complex). Exceptions were made for group items that used common tabular material, or where it was felt that one item might give away another if they were in close proximity. The arrangement approximates an arrangement by increasing difficulty. After 30 minutes, if any open-end tests had not been handed in, the experimenter cautioned the subjects they should soon proceed to the multiple-choice test. All formal instructions were presented by means of a tape recorder, except for answering any questions raised, seating students, and announcing the start and end of the break. The experimenter recorded the time when each test was handed in, so time spent working on each portion could be determined for each student. Most students completed both tests before the time limit of 55 minutes was up. (Mean total time on the Frequency Distribution test was 51.4 minutes, while mean total time on the Central Tendency test was 48.4 minutes.) As they were leaving, students were given a reminder card with the time and place of their next experimental session. The results will be described separately for the Frequency Distribution immediate test, the Central Tendency test, and the Frequency Distribution retention test. Copies of the materials used, and the instructions presented, are found in Appendix C.

1. Frequency Distribution Immediate Posttest. The data on this test were treated as a three-dimensional factorial design, with dimensions of Introduction, Presentation, and Sequence as shown below. The Sequence variable contrasts the A groups that took the Frequency Distribution test first with the B groups that had seen the Central Tendency film and taken the Central Tendency test the week before taking the Frequency Distribution test. The Central Tendency film was designed so that it could be shown to students who had not seen or studied any previous statistics. Since some of the central tendency objectives cover the relation of Central Tendency indices to various characteristics of frequency distributions, it was necessary to start this film with a rapid five-minute coverage of five objectives that were also covered in the frequency distribution film. Thus the B group had some previous

exposure to some of the Frequency Distribution objectives that might result in higher scores on the Frequency Distribution test. In particular, however, it seemed desirable to check the possibility that students would make effective use of the behavioral objectives only after they had had previous experience that indicated the test-to-come would actually be based upon those objectives. If this were the case, we might be expected to find that the P1 and I2 (and possibly I1) conditions were superior for the B group, but not for the A groups. It was necessary to randomly drop subjects from some groups to obtain proportional cell frequencies. The n's within cells in the first three-dimensional analyses are shown below.

Table 20

Frequency Distribution Immediate Post-Test

Sample Sizes Retained

Sequence

A groups (A1 and A2)
First Session
Introduction

Presentation	I1	I2	I3
P1	24	24	24
P2	30	30	30
P3	30	30	30

B group
Previous Session on C. T.
Introduction

	I1	I2	I3
P1	8	8	8
P2	10	10	10
P3	10	10	10

This same analysis was conducted on four response measures, Score and Time on the open-end test, and Score and Time on the multiple-choice test. The open-end test contained 21 items, 14 of which were fill-in items, requiring the insertion of one- or two-word responses in appropriate blanks. Other tasks included constructing frequency distribution tables and polygons and recognizing errors in printed tables and polygons. The multiple-choice test consisted of 38 items. Responses to these items were marked on IBM "port-a-punch" cards and were punched with a sponge and stylus when the test was completed, or at the end of the time period. All students finished the open-end test with no apparent time pressure. Some of the slower students, however, encountered time pressure on the multiple-choice test. Because of the relatively large number of analyses conducted in this experiment, in the body of the report we shall follow the practice of reporting only means that are pertinent to significant differences and, hence, represent more than merely chance variation. Similarly, only the F's of significant effects will be reported here, while the Appendix will contain the complete three-dimensional summary table, and the means, variances, and sample sizes of each cell.

2. Open-End Test Scores. All three main effects proved to be significant, while none of the interactions reached the .05 level.

Table 21

Analysis of Variance Data for Selected Open-End Test Scores,

Frequency Distribution Materials

Scores

Effect	F value	Prob.	Pertinent Means		
Sequence	14.970	p<.001	16.1 First session A groups		20.0 Previous session on C.T. B group
Presentation	3.075	p<.05	18.5 P1	17.1 P2	15.8 P3
Introduction	5.115	p<.01	18.2 I1	17.9 I2	15.1 I3

Appropriate t-tests were carried out on the Presentation and Introduction groups, using mean-square within under the assumption of homogeneous cell variances. The outcomes are indicated by Duncan's notation of a common line under means that are not significantly different (at the .05 level). Thus, under the Presentation condition, the only significant difference was that between the movie with objectives interpolated, P1, and the continuous movie, P3, (t of 2.479; other t's were 1.33 and 1.22). For the Introduction, however, the extraneous materials group, I3, was significantly inferior to both of the other groups (t's of 2.872 and 2.655 for I1 and I2 respectively), which did not differ from each other ($t < 1$).

3. Open-End Test Times. In this analysis, again, none of the interactions was significant. The main effects of Sequence and Introduction were significant, but the Presentation main effect was not.

Table 22

Analysis of Variance for Selected Open-End Test Times,

Frequency Distribution Materials

			Time in Minutes on Open-end Test		
Effect	**F value**	**Prob.**	Pertinent Means		
Sequence	22.809	$p<.001$	25.7 First session A groups		22.4 Previous session on C.T. B groups
Introduction	5.634	$p<.005$	23.8 I1	26.2 I2	24.6 I3

The group which received the behavioral objectives in the Introductory period, I2, took significantly longer than groups I1 and I3 (t's of 3.29 and 2.20 respectively), but the difference between I1 and I3 was not significant ($t = 1.08$).

Thus, on the open-end portion of the Frequency Distribution immediate posttest, we find that the B group scored higher and took less time than the A group, indicating superiority on both response measures. The P1 group scored higher than the P3 group, but there were no significant differences in time (means were 24.3 and 24.6 respectively). Thus, superiority was shown only on the score measure, but this superiority is clearly not due to taking more time. The Introduction effects are not as clear cut: both the I1 and I2 groups were superior in score to the I3 group. While the I1 superiority can not be attributed to greater time spent on the test, the I2 group did spend significantly more time on the test than the other two groups.

4. Multiple-Choice Test Scores. Only the main effect of Sequence was significant. The A group had a mean of 30.7, while the B group had a mean of 34.8 ($F = 8.09$, $p < .01$). Although the main effect mean scores fall in exactly the same order as they did on the open-end scores for the Presentation and Introduction conditions, the differences are not significant.

5. Multiple-Choice Test Times. Once more, the Sequence effect was significant, with the B group taking less time than the A group. The Introduction effects were also significant, with the means falling in the opposite direction from those on the Open-end time. The time spent by the I2 group was significantly less than that of the I1 and I3 groups (*t*'s of 3.85 and 1.99 respectively), but the I1 and I3 groups were not significantly different ($t = 1.86$).

Table 23

Analysis of Variance for Selected Multiple-Choice Test Times,

Frequency Distribution Materials

Effect	F value	Prob.	Pertinent Means		
Sequence	4.638	p<.05	26.8 A group		25.6 B group
Introduction	7.414	p<.005	27.7 I1	25.4 I2	26.6 I3

Apparently, the I2 group lingered over the open-end test, then speeded up on the multiple-choice test. An analysis of total times reveals only a significant Sequence effect. The total times of the I1, I2, and I3 groups are 51.5, 51.6, and 51.2 respectively, yielding an F of 0.173. The estimated standard deviation of the total time distribution is 5.4 minutes; it is clear that the distribution is negatively skewed, and this shorter time on the multiple-choice test for the I2 group may be a consequence of the fact that the slower students were terminated at the 55-minute time limit.

6. Central Tendency Immediate Post-tests. In this design, the A1 subjects who had completed the Frequency Distribution film and test the week before were continued under the same Presentation and Introduction conditions for the Central Tendency film. The B group subjects received this topic during their first session. Consequently, the B group now constitutes the group on their first presentation. To obtain proportional n's, cases were randomly dropped from cells, and the value of the cell mean was treated as a cell measure in three cells (but did not contribute towards the degrees of freedom within cells). The cell sizes used in the analysis are shown in the diagram below.

Table 24

Central Tendency Immediate Post-test Sample

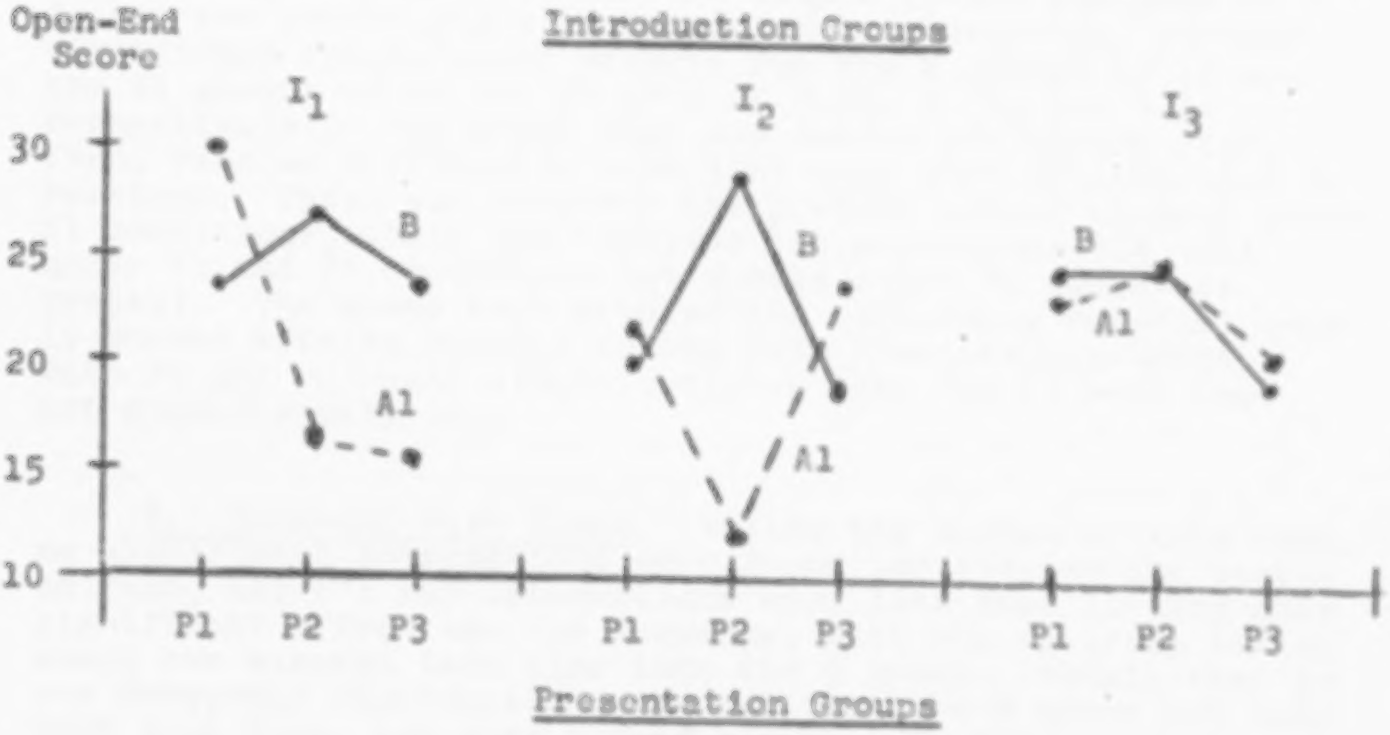
Sizes Retained

Sequence

	A1 Introduction groups			B Introduction groups		
Presentation	I1	I2	I3	I1	I2	I3
P1	6	6	6	18	18	18
P2	5	5	5	15	15	15
P3	4	4	4	12	12	12

7. Open-End Test Scores. This test consisted of 17 questions ranging from simple fill-ins to computing of medians and means to "explain" questions. The analysis of test scores contained two significant interaction terms, the main Sequence by Presentation interaction and the triple interaction term (F's of 5.966 and 3.359 respectively). Because the triple interaction term indicates the simple two-dimensional interactions differ from level to level of the third factor, simple

two-dimensional analyses were carried out on the Sequence by Presentation groups for each Introduction level (using ms within from all cells). The means are graphed below, with the significant terms of each two-dimensional analysis below the corresponding graph. (As usual, the three-dimensional summary table and cell mean and variance are reported in Appendix D).



Sequence x Presentation interaction, $F = 4.713$ $p < .025$	Sequence x Presentation interaction, $F = 7.869$ $p < .001$	Nothing Significant

A1 = Frequency Distribution Film First; Then Central Tendency Film

B = No Previous Film; Then Central Tendency Film

Figure 9. Graphic Presentation of Open-End Test Scores for Two-Dimensional Central Tendency Interactions.

The triple interaction is clearly associated with a lack of Sequence by Presentation interaction in the extraneous material Introduction groups (I_3 of Figure 9), a moderate interaction in the pretest Introduction groups (I_1 of Figure 9), and a strong interaction in the behavioral objectives Introduction group (I_2 of Figure 9). The difference between the B and A1 means for the P2 Presentation (dead air breaks) is significant for both the I1 and I2 condition (*t*'s of 2.236 and 4.182 respectively). The other A1 vs. B simple effects are not significant. Taking the simple random analyses of P in the two conditions with significant interactions yielded significant Presentation effects for the B groups of I2 and the A1 groups of I1 and I2 (F's of 5.93, 4.79, and 3.36 respectively). The group that was seeing its second film, then, reacted differently depending upon what introduction it received. Those who received the pretest scored highest under P1 conditions, those who received the objectives did well under P1 and P3 conditions but poorly under P2 (dead air breaks). The group that studied the extraneous material merely showed effects similar to the main Presentation means, with P1 and P2 means slightly higher than the P3 mean (but not significantly so).

8. Open-End Test Times. Unlike the scores of this test, no significant interactions were found for time on the test--in fact, all F's for interactions were less than 1. The only significant effect was for Sequence, with the A1 group taking about two minutes less time than the B group. Recall that in the Frequency Distribution open-end test the B group not only took less time, but also scored higher. In this case, the second-week group (A1) took less time but scored about two points lower (not significant).

Table 25

Analysis of Variance Data for Selected Open-End Test Times,

Central Tendency Materials

Effect	F value	Prob.	Pertinent Means	
Sequence	7.589	$p<.01$	15.1 A1 groups Second session	17.1 B groups First session

9. Multiple-Choice Test Scores. The test consisted of 39 multiple-choice questions, each of which had five possible responses. Unlike the Frequency Distribution test, there seemed to be no particular time pressure on this test, and all subjects finished. The mean total time for both sections was 44.0 minutes for the A1 groups and 49.6 for the B groups; however, as usual in academic tests, many subjects stayed the entire 55 minutes permitted (the standard deviation of total time measures was 6.22). The analysis of multiple choice scores from the Central Tendency test yielded no significant effects.

10. Multiple-Choice Test Times. The analysis of multiple-choice test times yielded a significant Presentation by Sequence interaction, plus significant main effects on both of these factors. The subanalyses that followed showed that both A1 and B groups show a trend across the Presentation conditions (F's of 6.424 and 7.14 respectively), but these trends are in opposite directions. The B group was fastest under P1 conditions, slower under P2, and slowest under P3. The A1 group was faster than the B group under P2 and P3, but about the same as B on P1. The means are shown below; all differences are significant except those for P1 and P2 for the B group and P2 and P3 for the A1 group. Taking the mean pairs vertically, the first pair is not significantly different, while the last two pairs are.

Table 26

Analysis of Variance Data for Selected Multiple-Choice Test Times, Central Tendency Materials

Effect	F value	Prob.	Pertinent Means		
Sequence by Presentation Interaction	9.576	$p < .001$	31.9 A1P1	25.7 A1P2	28.5 A1P3
			31.2 B P1	32.8 B P2	35.3 B P3

The lack of time pressure on this test as contrasted with the Frequency Distribution test may be noted in the mean time spent in the multiple-choice portions of each test. The group taking the test for the first time took an average of 26.8 minutes (A group) on the Frequency Distribution test, but an average of 32.8 minutes (B group) on the Central Tendency test. Similarly, the subjects in their second week took 25.6 minutes (group B) on the Frequency Distribution test, but 28.9 minutes (group A1) on the Central Tendency test.

11. Analyses Conducted on Only Those Subjects Who Had Completed All Tests Involved in Appropriate Comparisons. One of the problems of this experiment, like most experiments using volunteer subjects over a reasonably prolonged period of time, was drop-outs. All subjects were told to sign up only if they could come for three weeks in a row. They were also told they would received experimental credit in their Psychology 1 course only if they completed the experiment. This was in fact carried out, but there were still many subjects who completed week 1 but failed to show up for week 2, etc. In group B, only 78% of the subjects who completed the first week also completed the second week. In the A groups, about 73% of the first week subjects completed the retention test of the third week. The drop-out rate was somewhat increased by one week of flood conditions in the spring of 1964 which forced many students to move from their dormitories and double up in other dormitory areas. The scheduled experimental sessions continued without interruption, but the drop-out rate temporarily increased.

One possible explanation for the Sequence effects noted in the preceding analyses of the immediate test is that they are attributable to differential subject characteristics of drop-out students (who would be included in only the week 1 session) versus those students who did *not* drop out. To check this possibility, each of the Central Tendency and Frequency Distribution immediate test analyses was repeated (eight analyses in all) on subjects who had completed all sessions assigned, plus 14 A1 subjects who completed both the first and second week (same as the B subjects), but who failed to complete week 3. Any Sequence differences found could not then be attributed to drop-out effects, since no drop-outs could have occurred.

Again, it was necessary to randomly drop subjects from some cells to obtain proportional cell frequencies in the three-dimensional designs. The n's used are shown in the two figures that follow. It is important to note that even if a cell below has the same n as in the original analysis, the same subjects were not necessarily used in this analysis as in the original designs (unless no subjects were dropped in either design). In the original designs, subjects were randomly chosen from all subjects in that cell, whether they completed later sessions or not. This was desirable to maximize sample size and, hence, maximize power and the stability of the means reported. In the designs here reported, only no-drop-out subjects were used, and the subjects dropped were randomly chosen without any reference to their retention in the original analysis. Naturally, there is a high degree of overlap of subjects in the two analyses, and they may not be construed as independent statistical analyses. However, repeating the designs on no-drop-out subjects makes it possible to discredit this as a possible explanation for the reported Sequence effects and Sequence interactions.

Table 27

Frequency Distribution Immediate Post-Test,

Sample Sizes Retained

A groups
Introduction

Presentation	I1	I2	I3
P1	15	15	15
P2	12	12	12
P3	12	12	12

B groups
Introduction

	I1	I2	I3
P1	15	15	15
P2	12	12	12
P3	12	12	12

Total N = 234 (Total N in original analysis = 336)

Table 28

Central Tendency Immediate Post-Test

Sample Size Retained

A1 group
Introduction

Presentation	I1	I2	I3
P1	6	6	6
P2	5	5	5
P3	6	6	6

B groups
Introduction

	I1	I2	I3
P1	12	12	12
P2	10	10	10
P3	12	12	12

Total N = 153 (Total N in original analysis = 180)

In the eight three-dimensional analyses that were repeated, all of the terms that were significant in the original three-dimensional analyses were also significant in the no-drop-out analyses, with one exception. Conversely, none of the non-significant effects in the original three-dimensional analyses became significant in the no-drop-out analyses. These results discredit any interpretation attributing Sequence effects to drop-out effects, since the loss of power due to smaller n's would be expected to yield fewer significant effects. The only change was in the analyses of times on the open-end test of the Frequency Distribution immediate posttest. The Sequence by Introduction effect was significant in the original analysis (F = 5.634), but not in the no-drop-out analyses (F = 2.295). This particular interaction may be partially due to drop-out effects, or it may be that this change is simply the outcome of sampling fluctuation or the reduction of power. All other significant effects reported on in the preceding section stayed significant, with the means displaying the same trends. Naturally, the Means fluctuated somewhat, in this restricted sample, sometimes increasing the difference between means and sometimes decreasing it; the same is true of the various values of F. Since such fluctuations are inevitable, and since the outcomes of these analyses essentially duplicate those reported above but have less stability in the reported values of the means, they have not been included in this report. The sole consequence of these analyses, then, is to discredit the hypothesis that the significant Sequence effects obtained were due to drop-outs.

12. Frequency Distribution Retention Test. The Frequency Distribution retention test was taken two weeks after the students saw the Frequency Distribution film. The reminder cards told the students to report to a testing room run under the auspices of the Department of Psychology for administration of group vocational placement tests and academic make-up tests. Upon presentation of their reminder cards to the graduate assistant in this room, they were given copies of the two Frequency Distribution tests and instructed to complete the tests as before, turning in the open-end test prior to starting the multiple-choice test. Since the graduate assistants in this room were not hired by the experimenter, and since they had other duties, it was impossible to obtain accurate times on these tests or to accurately control the duration of the test. The subjects were told to try to complete the tests in a one-hour period, but slow subjects were allowed to continue beyond this period. The tests used were identical to those used in the first week, except for a note on the last page that told the subject that this concluded the experiment and thanked him for participating in it. The design contrasted

subjects run under the various Presentation and Introduction conditions, and the A1 and A2 groups, who differed in that that A1 group saw the Central Tendency film (and took the test) on week 2, while the A2 group had not seen this film or engaged in other experimental activity since the first week. Subjects were randomly dropped from cells to yield proportional n's for the Frequency Distribution retention test as follows.

Table 29

Frequency Distribution Retention Test, Sample Sizes Retained

Sequence

A1 Group

C.T. session intervening

Introduction

Presentation	I1	I2	I3
P1	6	6	6
P2	4	4	4
P3	9	9	9

A2 Group

No intervening session

Introduction

	I1	I2	I3
P1	6	6	6
P2	4	4	4
P3	9	9	9

13. Open-End Retention Test Scores. Two effects were significant in the three-dimensional analysis, the Sequence main effect, and the Presentation by Introduction interaction. Simple random analyses of variance on the three Presentation conditions for each Introduction group (collapsed over Sequence), and on the three Introduction conditions for each Presentation group, yielded only one significant F, that for the Introduction effect for the P3 group.

Table 30

Analysis of Variance of Certain Data from the Frequency Distribution, Open-End Retention Tests

Effect	F value	Prob.	Pertinent Means		
Sequence	13.544	p<.001	22.4 A1 group	15.9 A2 group	
Presentation x Intro. inter. (Subanalyses)	2.452	p<.05			
Introduction for P3 group	3.333	p<.05	20.3 I1	22.3 I2	14.5 I3

The A1 group showed an appreciable superiority over the A2 group. It is interesting to note that the original A mean on the immediate retention test was 16.1; the A1 mean, two weeks after taking this test and one week after the Central Tendency session, is 22.4, while the A2 mean is 15.9. Since none of the other simple Introduction or Presentation analyses was significant, *t* tests were conducted only on the means of the P3 group. The I2 and I3 means were significantly different (*t* = 2.48); the I1 and I3 difference was of dubious (*t* = 1.85) significance, while the I1 and I2 difference was not significant (*t* = .63). The trend of these means suggests that the P3 groups who had some contact with the objectives in the Introduction, either in the form of items in the pretest (I1) or in the objective list (I2), showed better retention than those who studied the extraneous material (I3). This same trend might have been expected for the P2 groups. However, neither the P2 nor P1 means show the same trend across the Introduction conditions, thus producing the significant I x P interaction. As mentioned, the null hypothesis was retained in all other comparisons.

14. Multiple-choice Retention Test Scores. Again the three-dimensional analysis yielded a significant Sequence main effect and a significant two-way interaction, but

this time the two-way interaction was Sequence by Introduction. The analyses of all Sequence comparisons for the three I conditions, and of the I comparisons for the two Sequence groups yielded only one significant difference, the A1 vs. A2 comparison for the I3 condition.

Table 31

Analysis of Variance of Certain Data from the Frequency Distribution Multiple-Choice Retention Tests

Effect	F value	Prob.	Pertinent Means	
Sequence (A1 vs. A2)	7.001	$p<.01$	37.0 A1 group	31.0 A2 group
Sequence by Intro. inter. (subanalyses)	3.529	$p<.05$		
Sequence for I3 group	10.654	$p<.005$	40.8 A1 group	27.9 A2 group

The original A mean on the multiple-choice test was 30.7. Again the A2 group is very close to the week 1 mean, while the A1 group has a significantly higher mean on the retention test. Unlike the open-end test, the multiple-choice test was taken under time pressure in the first week's administration, so not all subjects finished in the 55 minutes allowed. The administration of the retention test eliminated such time pressure, and all finished.

15. Control Groups.

Two control groups were run to clarify possible ambiguities in interpreting the experiment. The C1 control groups were treated like regular A subjects at the start of the week 1 hour, i.e., they were randomly assigned to the three Introduction conditions and worked on these for 12 minutes as usual. However, following this period, instead of seeing the

2

film and then taking the Frequency Distribution test, the subjects were released and told to report for a full two-hour session next week. In the following weeks, they completed the entire Central Tendency session, seeing the film under P1 conditions. The third week, they took the Frequency Distribution test. The C2 control group followed a similar schedule, except that in week 1, after studying the extraneous material (I3) for 12 minutes, they completed the Frequency Distribution test (in the usual 55-minute time limit) without ever seeing the film. In weeks two and three they completed the same schedule as the C1's, except that all C2 subjects were run under the I3 condition.

16. Frequency Distribution Immediate Retention Test. The C2 group (of 42 subjects) provided a measure of the general score level of Psychology 1 students in the absence of any specific instruction. The mean open-end score was 3.5, as contrasted to means of 16.1 for the A1 and 20.0 for the B groups. The difference between all of the regular groups and the C2 group is significant even by the inter-ocular test--i.e., hitting one between the eyes. The *t*'s for the above groups are 18.7 and 16.9 respectively, leaving little doubt that scores were appreciably raised by the filmed instruction. For the multiple-choice score, the C2 mean was 19.3 as contrasted with means of 30.7 and 34.8 (A and B respectively, with *t*'s of 9.0 and 9.5 respectively). Again there is no question that all groups learned. The mean C2 time was 25.9 on the open-end test and 20.9 on the multiple-choice test.

17. Central Tendency Test. The original predictions were that the A1 groups would have higher means on the Central Tendency tests. Increased familiarity with the Frequency Distribution concepts (that were thought to be necessary for some of the Central Tendency objectives) was expected to facilitate learning of several of the Central Tendency objectives and thus produce higher scores. The C1 control group was primarily designed to determine whether positive transfer was associated with experience on the introductory materials alone, and the C2 group also was to help determine how much might be attributed to the experience of taking the Frequency Distribution test but no Frequency Distribution film. However, the prediction was wrong--the scores showed no evidence of positive transfer from the Frequency Distribution session to the Central Tendency scores. In fact, the B mean was higher than the A1 mean on both the open-end and multiple-choice tests (not significant on either, however). It was also thought that if students benefited most from objectives only after completing a test that convinced them they were pertinent, then the C1 groups would show

Introduction effects like the B1P1 group, rather than the A1P1 group. The C1 means on this test are given below. Two factorial analyses in which A1P1, C1P1, and B1P1 constituted one dimension, and I_1, I_2, and I_3 constituted the other were conducted, but no significant effects were found on either open-end or multiple-choice scores. Again no evidence of positive transfer was obtained.

Table 32

Control Group Mean Scores on the Central Tendency Test

	C1			C2
C1 Open-end means:	$M_{I1} = 22.7$	$M_{I2} = 21.0$	$M_{I3} = 22.1$	$M_{I3} = 22.0$
C1 Multiple-choice:	$M_{I1} = 13.2$	$M_{I2} = 12.4$	$M_{I3} = 14.6$	$M_{I3} = 12.8$

18. Correlations Between Test Measures. The following correlations were run at the Michigan State Computing Center and are on fall, 1963, subjects who completed all assigned sessions. Only selected correlations are mentioned here; the complete tables are found in Appendix D. Introduction and Presentation conditions were ignored, leaving 32 A1 subjects, 77 A2 subjects, and 105 B subjects. The reliability of the Frequency Distribution test may be estimated by the test-retest correlations of groups A1 and A2. The open-end test showed correlations of .75 (A1) and .70 (A2), while the multiple-choice test showed correlations of .72 (A1) and .76 (A2). The time pressure existing during the Frequency Distribution test in the first week is shown by negative correlations between time on the open-end test and time on the multiple-choice test: -.70 (A1) and -.57 (A2). The B group, who had completed a session the week before, apparently did not experience this degree of time pressure, since the correlation was .07. The relative lack of such pressure on the Central Tendency test is shown by the correlations of -.21 (B) and .12 (A1) found on the two time measures.

The only direct data on the reliability of the Central Tendency tests was obtained by having two groups of B subjects report for a session on the third week, taking the same Central Tendency tests they had taken two weeks ago. The test-retest correlations obtained on these 16 subjects were .59 for the open-end and .39 for the multiple-choice test. These estimates, however, must be low due to sampling fluctuation. Under the assumptions of classical test theory, the maximum intercorrelation between measures of this reliability is .48 ($\sqrt{r_{xx} * r_{yy}}$; derivable from H. Gulliksen, *Theory of Mental Tests*, p. 101).[18A] Since the correlation between the Central Tendency open-end test and the Central Tendency multiple-choice test was .62 in group B and .70 in group A1, it is clear that the actual reliabilities must be somewhat higher than those from this very small sample. Other correlations of these tests with other test scores of the A1 group range from a low of .53 to a high of .73, again belying the low reliability value obtained in this small sample.

19. Frequency Distribution Retention Tests. The grand means of the C1 and the C2 groups are shown below; the difference between the groups was not significant in either case. In addition, the A1 and A2 means are listed for comparison.

Table 33

Mean Scores on the Frequency Distribution Retention Tests for

A1 and A2 Treatment Groups and C1 and C2 Control Groups

	C1	C2	A1	A2	Original C2 means
Open-end mean	10.8	16.0	22.4	15.9	3.5
Multiple-choice mean	30.1	31.4	37.0	31.0	19.3

The C1 open-end mean is significantly less than either the A1 or A2 means, but the C2 mean is significantly different only from the A1 mean. The control group means are not significantly different from the A1 or A2 means on the multiple-choice test. It is tempting to infer that the difference between the C2 group mean of 31.4 on the retention test and the 19.3 on

---

[18A] Gulliksen, H. *Theory of Mental Tests*. New York: Wiley, 1950, p. 101.

the original test is a measure of amount learned on Frequency Distribution subjects due to seeing the Central Tendency film (and taking the test). However, the fact that a 55-minute time limit prevented some subjects from completing the first test, but not the retention test, casts doubt on this interpretation, particularly for the multiple-choice test.

DISCUSSION

This section will be limited to the discussion of results and procedures that were either statistically significant or otherwise noteworthy for future investigations. The reader who wishes to review other details of this project can find them in the previous sections or in the Appendices of this report.

Interpolation of Information During "Breaks"

Most of the experiments in this series involved comparisons of materials used in "breaks" or interruptions of group presentations. A major experiment was conducted in which a continuous presentation was contrasted with presentations containing breaks, and the open-end test for the film used (on Frequency Distribution) showed achievement in the two conditions with "breaks" was significantly above that in the continuous presentation; the same trend was clearly shown in the Central Tendency open-end test, but the effect was not significant. This effect presumably results from increased audience attentiveness, as shown in Faison, Rose, and Pedell.[19] However, no effects were shown in the retention test two weeks later.

One clear virtue of breaks not shown in the experimental results but commented on by all instructors of the television presentations was that the breaks during which printed material was shown on the screen gave them a chance to relax. Knowing they were off camera, they straightened clothes and looked at notes while unobserved by the all-seeing camera eye. Many students in the experimenter's statistics class that received eight such presentations volunteered their appreciation of such breaks as a change of pace. Many studies in the literature involve the superiority of a condition in which interruptions occur over a continuous non-interrupted film. The material used in the interruptions varies from subtitles (Northrop, 1952),[20] spaced review (Miller and Levine, 1952),[21] or pupil participation questions (Lumsdaine, May, and Hadsell, 1958).[22] However, many such studies have failed to include "dead air breaks" to distinguish the effects of interruptions from the effects of the contents presented during the interruptions.

---

[19] Summarized in Lumsdaine, A. A. Student Response in Programmed Instruction: A Symposium. Washington, D. C.: National Academy of Sciences--National Research Council, Publ. No. 943, 1961.

From the viewpoint of both the instructor on television and the student, then, the use of such breaks seems desirable. The corresponding question concerns which material may be best used in such breaks, if any. This will undoubtedly vary with the nature of the objectives of the presentation. When the objective is mastery of a pure paired-associates list, such as a word response for each letter of the alphabet, evidence suggests that overt student responses to the specific stimuli be used.[23] When the stimulus and response terms are less specific, as in most academic contents, or when the task is not as clearly a paired-associate or rote memory task, the answer is not as clear.

In this study, unlike some in the past, or some recent programed learning studies, the position was adopted that the interpolated material should not be precisely the same material that later was used as the criterion measure. If a question were to be used in the presentation, answered, and then used in identical form as the criterion measure, this merely reaffirms the long-established fact that increasing trials produces increased learning. In addition, this procedure seems to make the implicit assumption that the specific set of questions used constitutes the complete population of all items that constitute the criterion behavior. A more reasonable position is that the specific questions used represent a sample from a large pool of possible items that measure the same objective. If a question can be used in the learning period that produces an increment on a different question in the criterion test, this may be interpreted as an indication of learning of a generalized objective rather than of a single specific question. As long as educators insist on the right to change wording in achievement tests, rather than using the exact wording of the text, usage of identical material as the "review" or other form of "independent variable" and as the criterion test should probably be avoided. The objectives used were, therefore, defined beforehand, and any questions written were regarded as specific instances of these objectives, as samples rather than as a population of criterion behaviors.

---

[20] Northrop, D. S. Effects on Learning of the Prominence of Organizational Outlines in Instructional Films. (Pennsylvania State Univer. Instructional Film Research Program) Port Washington, N. Y.: U.S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-33, 1952.

[21] Miller, J., & Levine, S. Summarized in Lumsdaine, 1961, op.cit.

[22] Lumsdaine, A. A., May, M. A., & Hadsell, R. S. In M. A. May & Lumsdaine, Learning from Films. New Haven, Conn.: Yale Univer. Press, 1958.

[23] Hovland, C. I., Lumsdaine, A. A., and Sheffield, F. D. Experiments on Mass Communication. Princeton, N.J. Princeton Univer. Press, 1949.

The experimental studies in the fall of 1963 failed to show any differences in student achievement when the students received the stem and answer of a question or responded to a multiple-choice question during the breaks. Providing the instructor with information on the students' responses also failed to show any effects, which is similar to the results of a previous study.[24] Of course, retaining a null hypothesis does not prove the equality of the means; in the statistics experiment in particular, the approach does not seem to be very precise. However, the experiment failed to show differences associated with active student response that have been found by Gropper and Lumsdaine.[25] This might be partly associated with the fact that the present experiments used multiple-choice responses and multiple-choice tests, whereas Gropper and Lumsdaine used fill-in blanks for responses in the interpolated material, and also used these extensively in the tests.

One of the liabilities of the use of such interpolated questions seemed to be that as specific instances of a more general objective, the students often seemed to attend to specifics of the question and miss the major point of the objective. In a sense, in the first studies, both the teachers and the "typical questions" served as an intermediary between the objectives and the student. The behavioral objectives were given to the teacher, and both he and the "typical questions" translated them into examples for the student. If the student had been exposed to the behavioral objectives directly in addition to the teacher's efforts, it seemed he might have learned more than when the objectives were communicated only in the indirect fashion. The experiments from this point on concentrated on variations in presentations of the objectives in the breaks.

The first significant result of these studies was a finding that the use of supplementary printed objectives during the breaks increased student achievement. Furthermore, the efficiency of these supplementary objectives was greatest if combined with a "dead air break" rather than with an oral presentation of the objectives during the break. For each of the three different speech contents, the combination of printed objectives to be used during dead air breaks produced the highest mean score. In a replication in the fall of 1963, with capitalized key words in the printed objectives, the combination of printed objectives with dead air breaks again yielded the highest mean achievement averaged over the three contents (although the interaction was not significant). Again the use of the supplementary printed objectives was clearly supported. The superiority of this condition may have been due to the fact that the subjects saw the objective in the same printed form as the questions on the test. In addition, the clear interruption of the audio presentation by the dead air period may have assured subjects' attendance to these printed objectives at appropriate times.

---

[24] Johnson, Games, Andersch, and Penson, op cit.

[25] Gropper, G. L., & Lumsdaine, A. A. The Use of Student Response to Improve Televised Instruction: An Overview. Report No. 7, Studies in Televised Instruction. Pittsburgh: American Institute for Research, Rept. No. Air-C13-61-FR-245 (VII), 1961.

The last statistics experiment failed to show such clear superiority of the use of interpolated objectives over plain dead air breaks when a film presentation was used with the objectives presented visually on the screen during the breaks. None of the comparisons of the means of the "dead air breaks" vs. "interpolated objectives" conditions was significant, and the relative positions of the means even changed from measure to measure. This change in relative effect of the interpolated objectives calls for an explanation, and only further experimentation can provide it. Possible hypotheses, however, might include the following.

1. The availability of the behavioral objectives *in printed form* during the speech presentations enabled students to use them at any time desired during the period, and this may at least have helped to produce the superiority observed. The crux seems to have been the availability of *printed* objectives rather than visual presentation of objectives (by slides during the breaks). The method of presentation may well have affected the subjects' orientation to the presentation in a desirable fashion; many subjects in the statistics films took notes, and the availability of printed objectives might have encouraged more continuous attention to the presentation. The speech experiment could well be repeated to test this hypothesis, with some groups receiving the objectives through slides during the breaks and others receiving the printed objectives as was done here. Similarly, printed-objective-groups could be added to the statistics experiment.

2. The effects of the use of objectives may depend on the type of responses called for by the objectives. As noted, the speech objectives came close to being paired associations of various short phrases or words. This is true of some

definitional objectives in the statistics topics also, but most of them do not seem to fit this rubric. It is quite likely that the efficiency of various techniques will vary with different content forms and that some form of content or task analysis is necessary before correct statements about the efficiency of various presentation methods may be made. We hope to rescore the statistics tests using only the items on objectives that are similar to paired-associates, to determine if the use of the objectives during the breaks produced an effect on these items that was masked by the other items.

3. The use of the introductory material in the statistics experiment may have created conditions in which the effect of interpolated objectives during the breaks was reduced. It was anticipated that the use of a pretest or the study of objectives during the introductory period might have an orienting effect similar to that produced by the interpolated objectives. This is one reason why the study of extraneous material during the introduction was used for some subjects. Such an effect should have produced a significant introduction by presentation interaction, but this was found only for the open-end retention test scores. All three kinds of introductory materials might, however, have induced a higher motivational set than was present in the speech experiment, and perhaps this interacted with the effect of the objectives.

It appears, then, that the use of interpolated material during breaks deserves further careful study. Certainly this method provided the most valuable leads to come out of this project. It is interesting to note the suggestion of the data for the desirability of "mixing media," i.e., providing students with printed material for use during "breaks."

Introduction Effects ("Set to Learn")

One aspect of study method that was investigated was the use of introductory materials. As mentioned earlier,[26] introductory materials may provide a generally strong "set to learn"; they may also serve a kind of "practice" or "instruction" purpose for the learner. In the relevant experiment here, the group that took the pretest and the group that received the set of behavioral objectives for study during the 12-minute introductory period of the statistics test were superior to the group that studied the extraneous material on the open-end test of the Frequency Distribution film. This is the only analysis where this trend was significant, although a similar trend was shown on the two multiple-choice posttests. All three sets of materials were judged to be relatively difficult for freshman students; this fact, combined with the fact that the students were instructed they would be tested after the film may have produced a strong "set to learn." Introductory pretests or instructions in previous studies have often directed attention to particular aspects of a film and have produced learning gains on this material, sometimes at the expense of learning on other material in the film. Since both the pretest and the list of objectives covered all of the objectives of the film, they may have served as an additional practice trial and/or they may have served to alert the student to the most pertinent portions of the film presentation.

---

[26] See Klare, op cit.

Further work on the effect of introductory material might well take the form of attempting to break effects down into components. A general "set to learn" effect might, for example, be separated from the specific "what to learn" effect. More operationally, several pretests of low, medium, and high difficulty ("set to learn") could constitute one dimension, while a second dimension might be the relevance of the test items to the objectives of the film ("what to learn" effect).

Use of College Classes in Experimental Work

One of the questions both educators and psychologists have been asking each other frequently of late is why more learning studies are not done in the classroom where their practical significance can be felt. The attempts to conduct the experiments in an ongoing statistics class, imposing the same variation in methodology on many presentations of different content to the students in the class, pointed out the extreme difficulties in this technique and provided at least part of the answer. The investigators were continuously busy in a rapid schedule of rehearsal, pretryout of test items on a graduate statistics course, preparation of criterion posttests, television production, supervision of test administration, supervision of test scoring, selection of items for insertion in regular course tests, etc., in addition to teaching the class as usual between the television productions. The necessity of using so many contents and so many different tests made it impossible to give as much attention to any single test as might have been possible if only one measuring instrument had been used. Administrative factors and design considerations made it impossible to break up the class into independent groups that could view different presentations of the same content using different methods. This, in turn, led to the use of the item as the sampling unit, with proportion of correct responses as the measure. This was probably very undesirable, since item characteristics probably determined this proportion more than variation in methodology of presentation or other "learning variables" did.

As a consequence of these liabilities, it was decided to conduct the remaining statistics studies on a limited number of topics, using volunteer students from General Psychology classes, with filmed presentations that have the great virtue of easy repeatability. It was felt that some of the advantages of in-class presentation might have been saved and some of the disadvantages avoided.

Sequence of Presentation

One of the astonishing conclusions to come from many programing studies has been the relative lack of effect of an orderly sequence vs. a random sequence of presentation. It is hard to believe that the careful sequencing done by generations of text-book writers has all been in vain, particularly in contents like statistics where there is rather substantial agreement that certain topics occur at the beginning of a course because they are needed for the later portions. Of course, it may be that some content areas involve no particular sequences and almost any random ordering would produce as much learning as any other. The three speech areas studied appear to lack any such clear sequencing. If behavioral objectives reduce to an approximation of independent paired-associates, there is no particular reason to expect any hierarchical sequence effects.

However, even if objectives do involve hierarchies, it may be that the hierarchies only cover relatively large blocks of material. Thus, even if studies are run of different sequences _within_ the large hierarchical blocks, sequence may make little difference. Within the Frequency Distribution and Central Tendency objectives, it was our opinion that a hierarchy probably could be found _within_ the Frequency Distribution objectives, but no evidence was found to support this in the summer, 1963, study. This, however, might be because the material had previously been assigned and presumably studied by these students; either scrambled or sequenced objectives may have been quite sufficient to orient the students to the behaviors desired, provide the stimulus and response terms of the paired-associates, etc. Further study should be done on naive students to test this hierarchy notion.

Within the Central Tendency objectives, there was little suggestion of any hierarchy; it seems, however, that many of the Central Tendency objectives were dependent upon mastery of Frequency Distribution concepts, so that this large block of material was hierarchically related to the Frequency Distribution material. Asking students to relate properties of the mode, median, and mean on skewed distributions certainly should be facilitated if they know what a skewed distribution is. About five of the 17 Central Tendency objectives covered in the test were thought to be dependent upon some degree of mastery of Frequency Distribution objectives.

From such considerations, the group that had the Frequency Distribution material previously was expected to be clearly superior to the group that did not on the Central Tendency test. This was not found at all; instead, positive transfer was found from the Central Tendency session to the Frequency Distribution session. This may be due to the fact that five Frequency Distribution objectives that were thought to be essential to any use of Frequency Distribution concepts were reviewed; in addition, many times frequency distributions and graphs were used as examples, with appropriate terminology. These additional examples, or these additional trials of the paired-associates, may have led to greater achievement on the Frequency Distribution test a week later.

The same phenomenon is found on the retention test. The group that saw the Central Tendency film in the intervening week was significantly higher on both retention tests than the group that did not. In short, none of the findings lend any support to the notion that the Central Tendency achievement was dependent upon prior mastery of Frequency Distribution objectives. Perhaps the cards were loaded in this direction by including the rapid run-through of the five Frequency Distribution objectives thought to be most crucial in the introduction of the Central Tendency film. Future studies will delete such sections when investigating possible hierarchical relations.

Another possible reason for a failure to show hierarchical relationships might be that the subjects were still in a very low stage of learning of Frequency Distribution concepts in all groups, and all subjects tended to miss the Central Tendency items based on the five hierarchical objectives. An item analysis to verify this has not been completed; however, if this were the case, we would generally expect relatively low Central Tendency test scores as contrasted to the Frequency Distribution test scores. However, comparing the composite mean score to the maximum possible score for both groups, we find ratios of .41 and .37 for the Frequency Distribution and Central Tendency multiple-choice tests, respectively, and of .34 and .37 for the Frequency Distribution and Central Tendency open-end tests. These values do not suggest that performance on Central Tendency was particularly low.

It is clear that sequence of presentation effects deserve further study. If the extra machinery of branching teaching machines or extra wasted pages of branching programs are to be justified, it must be shown that the students going through such controlled sequences learn more than when all students go through a single fixed sequence. Even more basically, if any of the various prescriptions for sequencing are to be taken seriously, it must be shown that a priori specifiable groups of subjects will learn more from a planned sequence than from various random sequences of the identical material.

Empirical studies of whether given behaviors are hierarchically related might clarify rules as to when careful (even individual) sequencing is needed and when it is not.

SUMMARY

For a summary of this set of experiments, see the Abstract presented at the front of the report.

APPENDICES

Appendix A. Behavioral Objectives

1) Objectives for Speech Presentations

2) Objectives for Statistics Presentations

The Frequency Distribution and Central Tendency objectives are those for the films in the fall of 1963. The objectives for the television presentations of fall, 1962, included these plus a few more, since there were two presentations.

Appendix B. Speech Materials

Copies of Speech Tests

Appendix C. Statistics Materials

1) Questionnaire Used in Statistics Experiment, Fall, 1962.

2) Instructions for Statistics Experiment, Fall, 1963.

3) Statistics Tests, Fall, 1963.

Appendix D. Complete Statistical Analyses, Statistics Experiment, Fall, 1963.

Supplement to the Partial Analyses Presented in the Body of the Report.

1) Frequency Distribution Immediate Post-Test, Analyses

2) Central Tendency Immediate Post-Test, Analyses

3) Frequency Distribution Retention Test Analyses

4) Correlations

APPENDIX A

Behavioral Objectives

1) Objectives for Speech Presentations

Note: The Speech objectives were used without capitalization in the experiment of spring, 1963. In the fall of 1964, some groups received them without capitalization, some received them with capitalization as typed below, and some received only the capitalized words. (In those objectives with capitals, capitals alone were not always in the precise order as in context.) See the write-up of the Fall, 1964, Speech Experiment for details.

Parliamentary Procedure

1. Every motion made during a meeting has a purpose.
2. When the purpose can be identified as one which PROPOSES ACTION, the motion is classified as a MAIN MOTION.
3. A SUBSTITUTE MOTION can be made which will CHANGE a sentence or paragraph of a MAIN MOTION.
4. A MOTION MAY BE DIVIDED into two or more main motions.
5. Given a main motion recognize a proper modification by substitution or division.
6. When the purpose of a motion can be identified as one which disposes of the proposed action by means OTHER THAN VOTING on the main motion that motion is classified as a SUBSIDIARY MOTION.
7. When the purpose of a motion can be identified as one which AFFECTS the PROCEDURE under which the meeting is conducted, that motion is classified as an INCIDENTAL MOTION.
8. When the purpose of a motion can be identified as one which MEETS EMERGENCIES, that motion is classified as a PRIVILEGED MOTION.
9. Given a specific motion, classify it according to its purpose.
10. Motions with the same purpose can be classified together according to this mutual purpose.

APPENDIX A (continued)

Propaganda

1. PROPAGANDA is an act of biased ADVOCACY in MASS COMMUNICATIONS.
2. INQUIRY is an act of ANALYZING controversial doctrines for the sake of promoting CRITICAL JUDGMENT.
3. CEREMONY is an act of intoning the sentiment-encrusted PHASES OF the creed in an assemblage of BELIEVERS.
4. INCITEMENT is an act of an impassioned harangue with a cry to VIOLENCE.
5. Given a statement and a context, classify it as either propaganda, inquiry, ceremony, or incitement.
6. The first condition for maximizing propaganda is monopolization of mass communications.
7. LACK OF OPPOSITION through the same or other mass media satisfies the condition of MONOPOLIZATION.
8. The second condition is CANALIZATION in which propaganda is maximized by associating the message appeal with already EXISTING AUDIENCE ATTITUDES.
9. The mass media are not effective in changing already existing audience attitudes.
10. Given a communication situation, identify the condition which maximizes its effectiveness.

APPENDIX A (continued)

Phonetics

1. ASSIMILATION is a SIMPLIFICATION OF the MOTIONS of the speech organs, a change in the direction of least resistance, greater simplicity, and greater ease within a given language in terms of neuro-muscular habits of the speaker of that language.
2. When one SOUND is CHANGED to a second sound under the influence of a THIRD SOUND, the process is called ASSIMILATION.
3. When a PRECEDING sound INFLUENCES the FOLLOWING sound, the process is called PROGRESSIVE ASSIMILATION.
4. Given a word in which the preceding sound influences the following sound, identify the process as progressive assimilation.
5. When the SECOND sound INFLUENCES the FIRST sound, the process is called REGRESSIVE ASSIMILATION.
6. Given a word in which the second sound influences the first sound, identify the process as regressive assimilation.
7. When two SOUNDS MODIFY EACH OTHER, both sounds disappear, and a third sound emerges in their place; the process is called RECIPROCAL ASSIMILATION.
8. Given a word in which the sounds influence each other, identify the process as reciprocal assimilation.
9. When the same sound or SIMILAR SOUNDS are changed or dropped to make PRONUNCIATION EASIER, the process is called DISSIMILATION.
10. Given a word in which the sound is changed or dropped, identify the process as dissimilation.

T.V. Research Project July 1, 1963

Frequency Distribution Presentation

Student Objectives

Note: "R&R" means "Recognize and Recall".

Introduction:

1. R&R: A set of numerical data (scores) can be represented as X: 1, 3, -3 -6, 8, 2, -13, 0, 4, meaning that X takes the values of 1, 3, -3, etc. "n" is the number of observations in the set of data. The set of data, X: -3, 2, 4, 0 has an n of 4. (Given a set of numerical data, symbolize all the individual score values by use of the symbol X. Symbolize the number of such scores by n.)

2. R&R: The purpose of a frequency distribution is to condense a set of data into more concise form, with little or no loss of information.

Definition and essential characteristics of a frequency distribution:

3. R&R: A frequency distribution is a record of the number of scores that fall in each score class, i.e., the frequency (f) of each measurement class (X). There are 2 elements: a) classes of observations or measures (X's).
b) frequencies in each class (f's).

4. Recognize violation of the fact that all observations must fall into one of the classes used and only one; not two. (Given tables which present groups of X's, identify which tables violate above.)

4a. In a frequency distribution meeting the above requirement, $\Sigma f = n$.

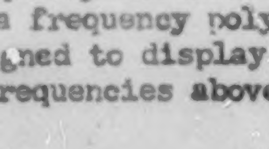
4b. Recognize $0 \leq f \leq n$ in discrete steps of 1.

5. R&R: Each score class between the extreme scores of the distribution should be listed, even if the frequency = 0.

6. Construct a frequency distribution table of interger scores (+ and/or 0 and/or - data). Make a table of the frequency with which X's appear in any specific group.

Graphing Frequency Distributions:

7. R&R: Frequency polygon always has f (frequency) on vertical axis and X (score value) on the horizontal axis. (Construct a graph which displays the X's on the horizontal axis and the frequency of X (f) as the vertical axis.



8. R&R: In a frequency polygon, plot the frequency above the midpoint of each interval, join the dots above the midpoints by a line, and bring the ends of the line down to a frequency of zero at the next class midpoint when it makes sense.

9. a. Convert a frequency distribution into a frequency polygon.
(b. Convert from a frequency polygon into a frequency distribution table.)
(On a graph designed to display the frequency of X's, from a given table plot the frequencies above the midpoint of the interval.)

Verbal Descriptions of Frequency Dist.

Strategy:

10S 1. R&R: Verbal definition of symmetric vs nonsymmetric (skewed). A symmetric graph may be folded so that the folded portion is superimposed over the unfolded.

10S 2. R&R: Verbal definition of unimodal vs multimodal. Unimodal is only one relatively high point on curve, multimodal is more than one.

10S 3. R&R: The above two dimensions are independent.

10a. Given a set of frequency polygons, identify the common form name for all in the set.

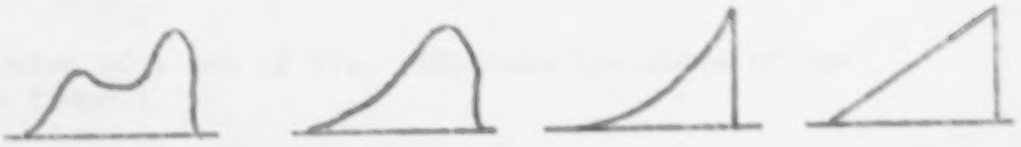
10b. Given a specific single frequency polygon, identify which form names are appropriate.

10c. Given a form name, draw or recognize a correct frequency polygon or histogram.

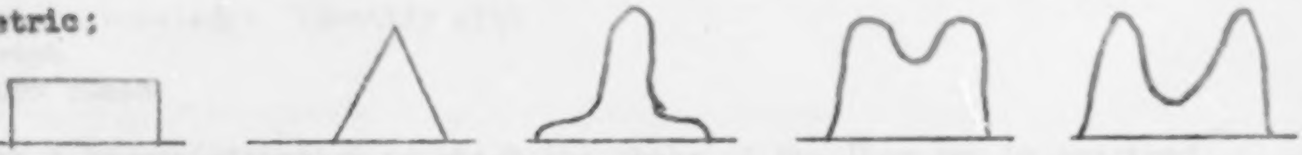
(Given a graph in which the points above the midpoints have been connected by a line, associate the shape of the line with the pattern it most closely resembles.)

Form names to be used above: negatively skewed, symmetric, positively skewed; unimodal, multimodal (bimodal); rectangular, U-shaped, J-shaped, bell-shaped.

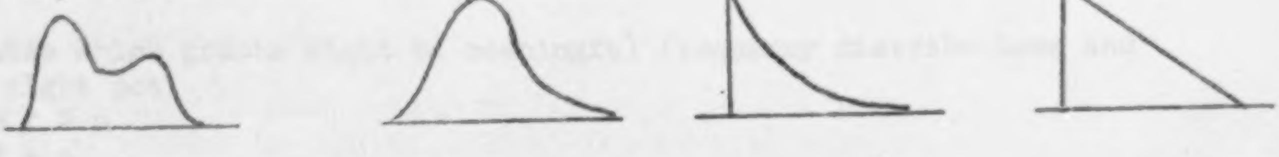
a. Negatively skewed:



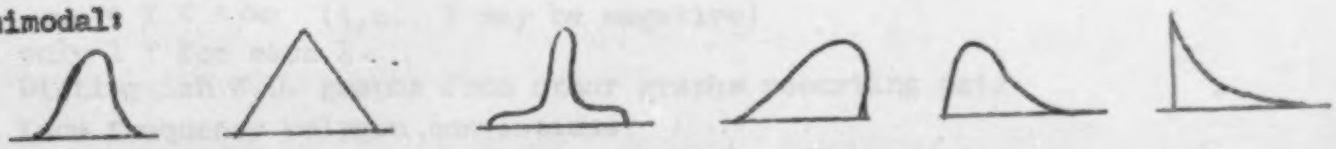
Symmetric:



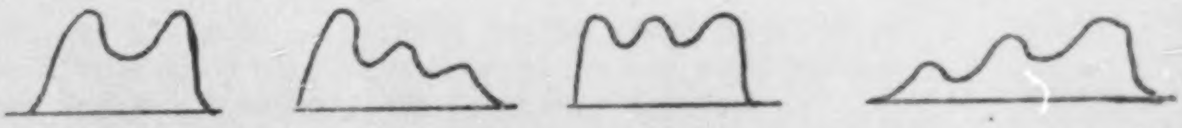
Positively skewed:



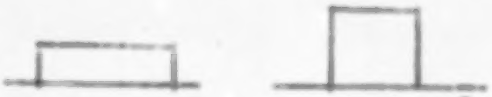
b. Unimodal:



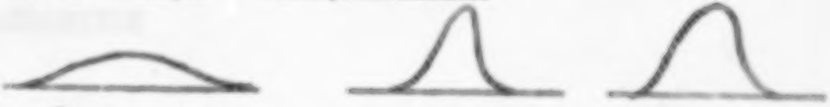
Multimodal:



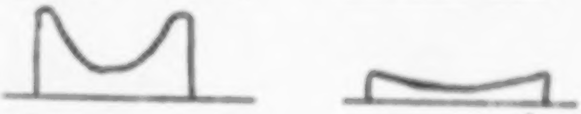
c. Specific names: rectangular:



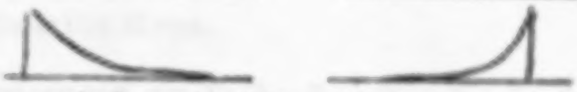
bell-shaped:



U-shaped:



J-shaped:



Strategy: Use verbal mediators of above.
Symmetric -- point of symmetry, fold over, etc.
Unimodal -- one high point

*11. Identify any frequency distribution table by application of above "form names."

12. Given a set of raw data, identify by form name.

Understanding Frequency Distributions and Applying.

*13. From a verbal description of the data, correctly identify data with
   a. graphs
   b. names

(From a word description of a set of X's, associate the shape of the line which would be found.)

*14. Given a common situation when shape of distribution is logically dictated by common knowledge, identify with
   a. graph
   b. form names

(Given a common situation in which the shape of the line can be imagined, associate the shape of the line which would be found.)

*15. Recognize which graphs might be meaningful frequency distributions and which might not.
   a. $0 \leq f \leq n$
   b. $\Sigma f = n$
   c. $-\infty < X < +\infty$ (i.e., X may be negative)
   d. only 1 f for each X.
   e. Distinguish F.D. graphs from other graphs reporting data.
   f. Know frequency polygon conventions.

T.V. Research Project August 12, 1963

STUDENT BEHAVIORAL OBJECTIVES

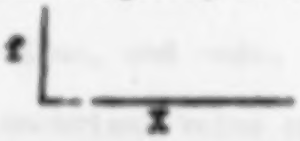
Central Tendency

A. Prerequisite objectives on Frequency Distributions.

1. A set of numerical data (scores) can be represented as X: 1, 3, -3, -6, 8, 2, -13, 0, 4, meaning that X takes the values of 1, 3, -3, etc. "n" is the number of observations in the set of data. The set of data, X: -3, 2, 4, 0 has an n of 4. (Given a set of numerical data, symbolize all the individual score values by use of the symbol X. Symbolize the number of such scores by n.

2. Construct a frequency distribution table of integer scores (+ and/or 0 and/or - data). Make a table of the frequency with which X's appear in any specific group.

3. Frequency polygon always has f (frequency) on vertical axis and X (score value) on the horizontal axis. Construct a graph which displays the X's on the horizontal axis and the frequency of X (f) as the vertical axis.



4. In a frequency polygon, plot the frequency above the midpoint of each interval, join the midpoints by a line, and bring the ends of the line down to a frequency of zero at the next class midpoint when it makes sense.

5. 5a. Convert a frequency distribution into a frequency polygon.
   5b. Convert from a frequency polygon into a frequency distribution table.

B. Central Tendency Objectives Covered.

6. Recognize: Central Tendency is description of "average" — a single point on a score scale to represent the set of data — the central location of data.
   Distinguish between 2 distributions grossly different in Central Tendency; identify the larger distribution, etc.

7. Recognize: Central Tendency and "average" or "location" are loose general descriptions, while mean, median and mode are specific indices.

Mode

8. Recognize verbal definition: The mode is a point on the score scale which has a frequency that is larger than other frequencies in its neighborhood.
   a) Recognize: Modes are score values, not frequency values.
   b) More than one mode is possible in a distribution.

. Identify modes from graphs and tables.
Distinguish unimodal, bimodal, and multimodal frequency distributions:
a) graphs, b) tables.

edian

). Recognize the verbal definition: The median is the point on the score scale which has 50% of the cases above it and 50% below it.

1. Compute the median on a short set of ordered or unordered integral data with the median falling at or between successive integers.

ean

2. Recognize the verbal definition: The mean is the point on the score scale corresponding to the sum of the scores divided by their number.

$$\bar{X} = \frac{\Sigma X}{n}$$

). Compute mean on any short set of integral data (4 or - numbers) or any meaningful frequency distribution or graph representing such a set of numbers.

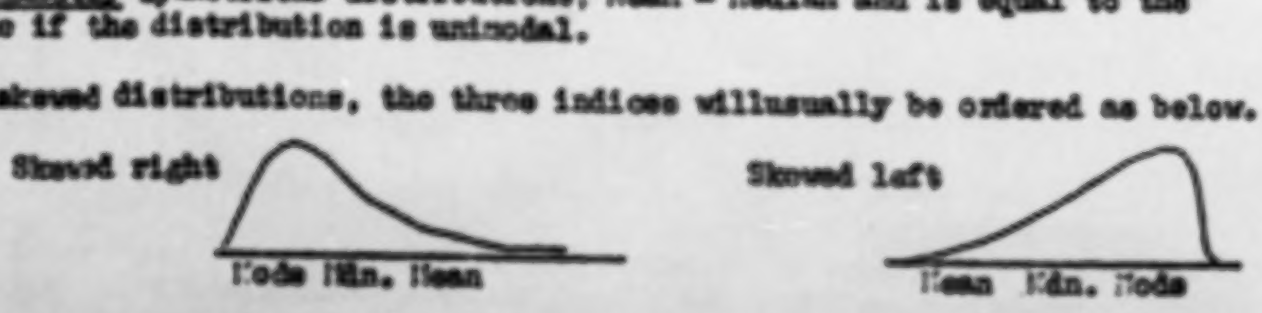
mparison of properties of mean, median, and mode.

. Only the mean is based on the numerical value of every score, because only the mean is based on the total. (X).

5. Recognize when change in score value will affect mean, median, and/or mode or how added score values will affect each.

7. *In __exactly__ symmetrical distributions*, Mean = Median and is equal to the mode if the distribution is unimodal.

3. In skewed distributions, the three indices willusually be ordered as below.



). Recognize: the mean is affected by the presence of a few extreme scores while the median and the mode are not.

lection of appropriate average for various questions.

). Recognize: Prefer mean if
a. approximately symmetrical distribution __or__ if
b. interest is centered on numerical value of all scores, i.e., total rather than typical
(e.g., total effect of betting on roulette wheel over long run — the mean.)

21. Recognize: Prefer median if interested in typical rather than total
and a. distribution is skewed, e.g., a few extreme values or
b. values are missing so can't compute mean.

22. Recognize: prefer mode only if multimodal situation and multivalued index is usable.

T.V. Research Project June 25, 1962

BEHAVIORAL OBJECTIVES

"Spread or Variability of a Distribution"

(items 34-50 are also test items or modifications of these)

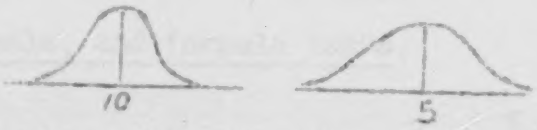
I. A general concept of Variability

* 1. Distinguish between 2 distributions grossly different in variation.
  - a) Numerical data
  - b) Graphs

* 2. Distinguish between central tendency and variability.
  - a) R & R: C.T. is a point on a score scale.
  - b) R & R: Variability is usually expressed as a distance on a score scale.
  - c) Classify C.T. and Variability indices as point or distance measures.
  - d) Recognize measures of variability as such.

asure of deviation;
tent to which scores
viate from C.T.

* 3. Recognize the independence of C.T. and variability.
  - a) Graphs



which has greater C.T.
which has greater Variability

  - b) Verbalize, etc., that the C.T. may be large and the variance small, and vice versa, or equal C.T.'s may have different variances.

II. Express Variability as a quantitative measure.

A. Range.

* 4. Verbally define range (R & R)

* 5. Compute range from ordered or unordered (+ or -) numbers.

6. Order by size of range graphical or numerical data, i.e., recognize relative size of range of any set of data.

7. R & R: Disadvantage of range being based on only 2 measures of a distribution. (the extreme scores are on either end, therefore very unstable.)

B. Variance and Standard Deviation -- generally.

8. R & R: These indices are based on numerical value of all scores.
  - a) R & R; If change value of any score, that changes these.
  - b) Distinguish whether similar to or different from other indices on this property. (Classify)

* 9. Recognize: Variance, S.D. and range are specific indices of variability. Variability is a loosely used word for a general characteristic. Variance and Variability are not synonymous.

  - c) Variance and S.D. show less sampling variability than range since based on all cases.

C. Population Variance = $\sigma^2 = \frac{\Sigma \chi^2}{N} = \frac{\Sigma (X - M_X)^2}{N}$

(I) Distinction between population and sample:/review topics pertinent here.

(review)
10. * R & R: between M_X (Pop. mean) and $\bar{X}$
(review)
11. * R & R: between N (Population size) and n (sample size).
(review)
12. * R & R: Deviation score (deviation from mean):

$x = X - M_X$ for population or $u = U - M_U$, etc.

$x = X - \bar{X}$ for sample .

(review)
13. * Recognize: Both cases (above) are covered by the same symbol.

(Note to instructor: please write x as a script small "X" -- i.e., as χ when using blackboard.)

(II) New name, symbols, and formula tasks.

* 14. R & R: Population variance = $\sigma_X^2 = \frac{\Sigma \chi^2}{N} = \frac{\Sigma (X - M_X)^2}{N}$.

$\sigma_V^2 = \frac{\Sigma v^2}{N} = \frac{\Sigma (V - M_V)^2}{N}$.

a) e.g., Population Variance is symbolized as ________

b) Distinguish from incorrect formulas, e.g.:

What is correct: $\sigma^2 = \frac{\Sigma (X - \bar{X})^2}{n}$ etc., or $\frac{\Sigma (X - M_X)^3}{N}$?

c) R & R: Express formula verbally.

* 15. R & R: σ^2 is a parameter.

* 16. Compute Variance by this formula on short (+ or -) sets of population data with integral means. (different X's and N's; ordered and unordered data). (or such data represented by graph)

* 17. R & R: $\Sigma \chi^2 \geq 0$, therefore $\sigma^2 \geq 0$, i.e., never negative.

a) Only 0 if all X's have the same value -- otherwise +.

b) Recognize $-\sigma^2$ as impossible.

D. Standard Deviation is Square Root of Variance (Var. = sq. of S.D.)

* 18. R & R: S.D. = $\sqrt{\text{Variance}}$; Variance = (Standard Deviation)^2

a) Symbols: S.D. is σ, the Variance = ?

b) Numbers: Variance = 9, S.D. = ?

c) Formula

d) Words

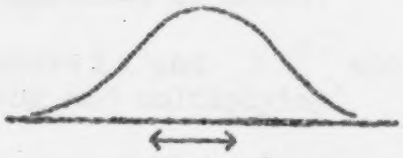
e) Above holds for population and sample.

* 19. Compute S.D. from short sets of data (as in 16)
a) assume can use tables to solve for square root.

20. R & R: S.D. is in original raw score units (e.g. lbs., etc.) while Variance is in squared units.

Therefore: S.D. can be graphically represented in a frequency distribution while Variance cannot. Because of this, the S.D. is most widely used index of variability.

a) Recognize:



σ or σ^2 ?

b) Recognize S.D. is in same units as range, mean, median, etc. -- Variance is not. (S.D. range unless both zero)

c) S.D. is distance on score scale.
Variance is squared distance.

E. Variance is mean of the squared deviation scores.

21. Recognize $\sigma^2 = \frac{\sum x^2}{N}$? = Mean of x^2 = M_{x^2}

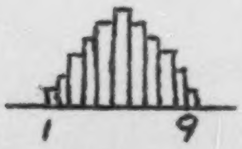
a) in symbols
b) in words.
c) in data form

22. R & R: Properties of means apply.
a) Therefore, large x^2's increase σ^2.
b) Small x^2's decrease σ^2.

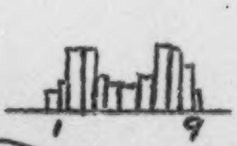
* 23. Size of variance and S.D. is independent of number of cases.

* 24. Determine by inspection of graphs which data will have larger σ^2's (S.D.s)
a) Same range and n.
b) Different range and n.

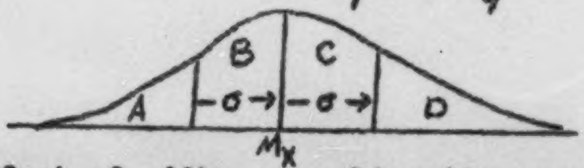
* 25. Effect on σ^2 of changing distribution by adding or subtracting scores.



changes to



etc.



Effect of adding or subtracting scores from A, B, C, & D on M_X and σ_X.

F. Computational Formulas.

(I) Computation of Σx^2 by use of raw data (X) rather than deviation data (x).

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = \Sigma X^2 - N\bar{X}^2 = \text{Sum of Squares}$$

*26. Use on short + or - data (regardless of mean).

* 27. Use on longer sets of data where ΣX and ΣX^2 and ~~M~~ N is given (avoid spending time in adding and multiplying).

* 28. Recognize it yields the same numerical answer as Σx^2 computed directly.

* 29. Recognize it is more convenient if $\bar{X}$ is not an integer. (Which is more convenient on a given set of data?)

≠ 30. Recognize Σx^2 is called the sum of squares.

BEHAVIORAL OBJECTIVES

VII. Normal Curve

I. Comments, Definition and Description

1. Recognize: Normal curve is a mathematical model defined by an equation.

   Strategy: ship models in tank --
   math model is

$$y = f(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-(X-M_X)^2/2\sigma^2}$$

2. R & R: (minor points) Normal Curve model is a particular symmetric bell-shaped curve with
   a) maximum height at mean ($M_z = 0$)
   b) inflection points at ± 1 σ from M_X (z's of ± 1)
   c) tails asymptotic to zero.
   Therefore, theoretically, X extends to ± infinity, but practically, over 99% of the cases fall within ± 3 σ's.

3. R & R: Normal curve is a form; may have many normal curves with different means and standard deviations.
   Many N.C.'s -- not just "THE" N.C.

   Strategy: Normal curve is defined by an equation in which M_X and σ_X are in the equation and determine the C.T. and Var. of the curve. The equation itself determines the form.

   a) Recognize: N.C.'s may look somewhat different due to different σ's and different units on axes. (T from F.D. 18)

4. Recognize: Other symmetric, bell-whaped curves may still not be normal -- may be too flat (platykurtic) or two peaked (leptokurtic).

   Strategy: Greek:
   platy means broad, wide or flat.
   kyrtos means curved.
   leptokurtic means finely or thinly curved.

   a) Distinguish N.C. graphs from grpahs of distributions that are grossly leptokurtic or platykurtic, or skewed or multimodal, etc. (T from FD 20 + 21)

5. Recognize: Models (mathematical and otherwise) are useful to the extent to which they approximate the real world -- normal curve is very useful because it is often a very good approximation of certain frequency distributions. (T from F.D. 20, 22, 23)

   a) Recognize: Many kinds of scores (data) are not N.D., i.e. can't apply N.C. model automatically.

   Strategy: Age data in all U.S. male widowers, etc.
   Bust size at Girl Scout picnic with Hyllywood starlets as hostesses.

b) Recognize: none of the following N.C. table solutions will work if the distribution is not normal.

6. Recognize: Types of data that are usually approximately normally distributed.

a) Recognize: if raw data is a collection of chance (random) events, it is usually N.D.
i.e. coin flips, errors of observation.

Strategy: History -- errors of observation in astronomy and physical sciences.

b) Scores of widely used objective type tests are usually N.D. when applied to population for which designed.

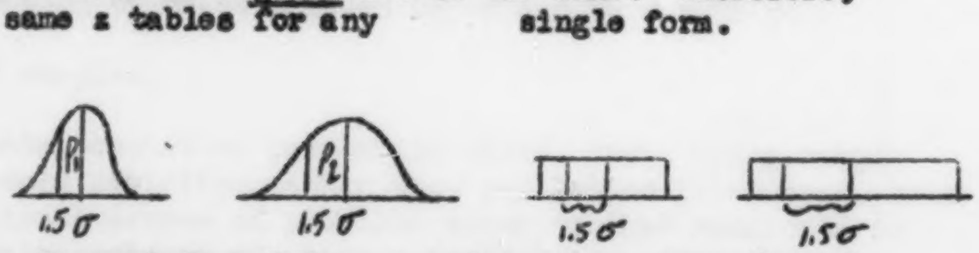
7. Recognize: Normal curve model is a model of frequency distributions of populations of continuous data. Therefore, use on discrete recorded data is always an approximation.

Strategy: Histogram with superimposed N.C.

8. R & R: Because normal curve is a continuous curve, frequency or proportion must be determined from area rather than height.

9. Recognize: for any given exact form, the p and f and % of area under the curve between 2 given z's is the same. Therefore, can use the same z tables for any single form.

$P_1 = P_2$



Review: Convert from X to z or vice versa given formula:

$$z = \frac{X - M_X}{\sigma_X}$$ to convert X to z

($M_z = 0$, $\sigma_z = 1$ for any complete z distribution)

$X = z\,\sigma_z + M_X$ to convert from z to X.

10. R & R: To simplify model of N.C. so won't have different values in the equation (and therefore different tables) for each different M_X and σ_X, use z-scores.

Therefore, we always have the same formula and the same tables for any situation.

Strategy: $$y = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} = C\, e^{-\frac{1}{2}z^2}$$

a) Given a normal distribution with specified M_X and σ_X, we can determine z scores and therefore we can determine percentiles and percentile ranks of scores from tables with no further information needed.

II. Use of Normal Curve tables on N.D. populations
(Table 5.1, found on page 60 of Guest)

11. Use table to determine Percentile Rank of z from any N.D., or of X from any N.D. with given M_X and σ_X.

    a) Recognize: verbal question as being of this form.

12. Use table to determine Percentiles (in z units) given %, or in X units given M_X and σ_X of a N.D.

    a) Recognize verbal question as being of this form.

13. Use table to solve for % or p between any two z's or any two X's, given M_X, σ_X of a N.D.

    a) Recognize verbal question as being of this form.

14. Given z (or x and σ_X) or N.D., use tables to solve for the p of cases between + z and -z (or between $M_X + x$ and $M_X - x$).

15. Recognize: Use tables to solve for x or z to include (or exclude) a given p (or %) closest to the mean.
    i.e., Recognize: given p and σ_X, solve for x such that the proportion of cases between $M_X - x$ and $M_X + x$ is p.

III. Populations and samples.

If must estimate population percentile ranks, etc., using sample mean and standard deviation rather than population parameters, this adds an additional source of possible error to that mentioned in obj. 7, and makes outcome of above a decided approximation.

Strategy: reporting percentile ranks beyond two decimal places is usually rather absurd.
62.37 %, etc.

Course ____________________
Instructor ____________________
Semester ____________________

Name of Student ____________________

Date ____________________

Most courses have a required attendance period--usually a 3 hour course has 3 hours of required attendance each week or approximately 42 hours a semester. This time may be spent in many ways, and on various aspects of the instructional process. The instructor must decide how he can best use the time to help the student to learn. I would like your judgments and opinions on how this can be done. Any comments you make here will in no way affect your grade.

Please follow the directions given in the 9 numbered steps.

Step 1. (Class time)

I have arbitrarily divided the instructional process into 5 aspects indicated by the Roman numerals. Divide the total class time (100%) into the percentages you would like to see devoted to each of the 5 aspects; use the blanks on the left. Look at the five aspects carefully--do this before you continue below. Zero % is a permissable response if you feel no time should be spent on one aspect, but make sure the total equals 100%.

_____% I. Communicating course objectives to the student (indicating what behaviors are desired).

_____ _____ _____ _____ _____ _____ _____ _____

_____% II. Directing student behavior so the student acquires the course objectives (teaching the desired behaviors).

_____ _____ _____ _____ _____ _____ _____ _____

_____% III. Informing the student of the extent to which a specific objective is acquired or whether further work on the objective is necessary (feedback to student of adequacy of learning to date).

_____ _____ _____ _____ _____ _____ _____ _____

_____% IV. Clarifying ambiguities in I, II, & III above and correcting the misconceptions of individuals (answering student questions).

_____ _____ _____ _____ _____ _____ _____ _____

_____% V. Measuring and evaluating the degree of individual and class acquisition of the course objectives (feedback to the instructor on the adequacy of the course work in meeting the objectives and student evaluation for grading purposes).

_____ _____ _____ _____ _____ _____ _____ _____

Step 2. (Class time)

Now look over all of the material on page 4, parts A and B. Do not continue until you have looked over all alternatives.

Step 3. (Class Time)

Now use the 10 sections of Page 4, A, to indicate which techniques you would use on aspects I to V above by listing them in the blanks below each aspect. Choose the methods (1-10) you would use for each aspect that you gave a % greater than zero. One or more techniques may be used on each aspect, and the same technique may be listed under several aspects if you feel the technique can be used to accomplish several aspects. List the techniques in order of importance, if you use several under any aspect.

Step 4. (Class Time)

List the % of the total class time that you would use with each technique (1-10) in the space below.

(1) _____% (2) _____% (3) _____% (4) _____% (5) _____%

(6) _____% (7) _____% (8) _____% (9) _____% (10) _____%

Zero % is again a permissible response; be sure that the sum = 100%.

Step 5. (Materials for Free Time)

Indicate what you would like to have available for use on your own time. Divide the total time of your free study into percentages under the 3 aspects below.

_____% I. The communication of course objectives to the student (the indication of what behaviors are desired).

_____ _____ _____ _____ _____ _____ _____ _____

_____% II. The acquisition of course objectives by the student (the learning of desired behaviors).

_____ _____ _____ _____ _____ _____ _____ _____

_____% III. The verification by the student of the extent to which a specific objective is now acquired (Feedback to the student on the adequacy of learning to date).

_____ _____ _____ _____ _____ _____ _____ _____

Step 6. (Materials for Free Time)

List the materials by number from page 4, B, that you feel would be important. Put them in order of importance in the blanks below each aspect. (Note: if you did not use #1 A on the first page, you cannot use #1 B; since without a T. V. lecture you would have no T. V. notes.)

Step 7. (Materials for free time)

Again, list the % of your free time you would expect to spend on each of the above materials.

(1) _____% (2) _____% (3) _____% (4) _____% (5) _____%

(6) _____% (7) _____% (8) _____% (9) _____% (10) _____%

Zero % is again a permissible response; check that the sum = 100%.

Step 8. (Evaluation)

The instructor must check that course objectives are acquired and he must evaluate the student. Assume the amount of class time for this purpose to be approximately the % you indicated for area V. Now, indicate the % of your final grade you would prefer to have dependent on each of the following techniques. Indicate this in the blanks on the left.

_____ 1. Closed book test items--multiple choice or matching. _____

_____ 2. Closed book test items--essay and definitions. _____

_____ 3. Open book test items--computational and methodological. _____

_____ 4. Pop quizzes--short, unannounced; given occasionally on days reading assignments are supposed to be finished. _____

_____ 5. Graded overnight exercises. _____

_____ 6. Graded exercises done in class under graduate student supervision. _____

Once again zero % is permissible; be sure your total equals 100%.

Step 9.

Now, in the blanks on the right, put the % of the time for each choice that you feel would produce the most desirable student behavior, and the greatest amount of learning. Ignore your own preferences, and put the figures you would use if you were the instructor, and wished to maximize student learning. Again, zero % is permissible, but be sure your total equals 100%.

This page may be removed and used for reference when completing the other pages. Complete the pages in the order indicated. This page need not be turned in; the others must be.

Thank you.

A. Instructional Techniques and Materials (for portion of course that requires attendance).

1. T. V. or film presentation
2. Large lecture (250 students), no student questions possible.
3. Small lecture (40 students), student questions possible and encouraged.
4. Class discussion (40 students). Instructor will answer questions only, no lecturing beyond particular points raised.
5. Small group discussion (10 students) by graduate student.
6. Use of small step program such as TMI statistics program for time equal to the above.
7. Use of large step program such as Games' Sampling Distribution of the Mean for time equal to the above.
8. Doing exercises under graduate student supervision (groups of 10) for equal time, with answers given in the period.
9. Taking tests.
10. Going over previous tests.

B. Free Access Material (for portion of course outside of class hours, with no compulsory checking of student usage).

1. Your own personal notes from T. V. or Film presentation.
2. Your own personal notes from Large Lecture presentation.
3. Your own personal notes from Small Lecture presentation.
4. Your own personal notes from Class Discussion.
5. Your own personal notes from Small Group discussion.
6. Small step program, in book form.
7. Small step program, available in machines in library reserve room or service room such as Porter 211.
8. Large step program, in book form.
9. Large step program, available in machines in library reserve room or service room such as Porter 211.
10. Exercises with answers provided.
11. Exercises to be completed, handed in, corrected, and returned one week later.
12. A printed list of course objectives on each topic (such as received on half of the T. V. presentations).

NOTE: Please choose only one of items 13, 14, and 15.

13. A large expository text book--say 230 pages on descriptive statistics and 130 pages on hypothesis testing and estimation (for a total of 350 pages--like Blommers and Lindquist).
14. A small reference text--both more concise and more complete than Guest--about 130 pages total (like Edwards' Statistical Analysis).
15. An intermediate text--about 200 pages total, in between above two (like Guest in amount of exposition).

Go back to page 1, Step 3, now.

APPENDIX C

Instructions on Tape Recorder. Experiment F63: Statistics

Week I--Groups A1, A2, and B.

(Start)

This experiment is on descriptive statistics. This is a topic in your psychology 1 course content. We do not expect you to know much about descriptive statistics now, but we hope that you will know more when you finish the experiment. What you learn during this experiment may help you during your final exam in psychology 1. Today we're going to show you a film. Pay attention to the film and learn as much from it as you can. After the film is over, you will be tested on the material covered in the film. Before you see the film, however, we will spend some time working on the materials you just received. You will have twelve minutes to spend on this material. Start working on the material you have now. We will stop you in twelve minutes. If you have any questions, raise your hand and the experimenter will see you.

(After the film and rest)

We will now take a test on the material covered in the film. Leave the test face down until I tell you to start. The test is in two parts. The first part is completion and short-answer questions; the second part is multiple-choice. You will not need the sponge and black plastic stylus in the first part, so lay the sponge and black stylus on the floor under your chair. You will have 55 minutes on this test. Try to answer every question. Do not spend too much time on any one question. If you are not sure of the answer, write down your best guess and continue to the next question. When you complete the top test, the short-answer test, raise it in your hand and the experimenter will come around and take it from you. As soon as he has taken the first test from you, turn over the multiple-choice test, read the instructions and proceed to answer the multiple-choice questions. To repeat, you have a total of 55 minutes for both tests. Do not spend too much time on one item; keep moving. As soon as you have completed the short-answer test, raise it in your hand. As soon as the experimenter has taken your first test, continue on to the multiple-choice test. When you are finished with both tests, you may leave. Do your best, but try to finish all items in both tests. If you have any questions, raise your hand after we start. Ready, turn the top test over and start.

APPENDIX C (continued)

Week 2--Groups A1 and B.

(Start)
Our procedure today will be same as last week. First, you will work on the printed material you have; then we will see a film on a new topic. After the film we will have a test on this new material. Start working on the material you have now. We will stop you in twelve minutes. If you have any questions, raise your hand and the experimenter will see you.

(After the film and rest)
Now we will take a test on the material covered in the film. Leave the test face down until I tell you to start. The test is in two parts. The first part is completion and short-answer questions; the second part is multiple-choice. Place the sponge and black plastic stylus on the floor until you need them. You have a total of 55 minutes for both parts. Do not spend too much time on one item; keep moving. When you have finished the short-answer test, raise your hand and the experimenter will come around and take it from you. As soon as he has taken the first test from you, turn over the multiple-choice test, read the instructions and proceed to answer the multiple-choice questions. When you are finished with both tests, you may leave. Do your best, but try to finish all items in both tests. If you have any questions, raise your hand after we start. Ready, turn the top test over and start.

Week 1--Group C1.

(Start)
This experiment is on Descriptive Statistics. This is a topic in your psychology 1 course content. We do not expect you to know much about Descriptive Statistics now, but we hope that you will know more when you finish the experiment. What you learn during this experiment may help you during your final exam in psychology 1. Today, you are going to spend some time working on the material you just received. You will have twelve minutes to spend on this material. Start working on the material you have now; we will stop you in twelve minutes. If you have any questions, raise your hand, and the experimenter will see you.

(After the twelve minutes)
This is all we will ask you to do today. Thank you for coming in. We apologize for keeping you such a short time this session, but this is necessary in the experiment. Next week, when you come in at the same time, you will see a film and take a test on the film. Next week, you will have a full

APPENDIX C (continued)

two-hour session--keep your reminder card until then, and be sure to show up on time for a full two-hour session next week. That is all for this time. Thank you.

Week 1--Group C2.

(Start)
This experiment is on Descriptive Statistics. This is a topic in your psychology 1 course content. We do not expect you to know much about Descriptive Statistics now, but we hope that you will know more when you finish the experiment. What you learn during this experiment may help you during your final exam in psychology 1. Today, you are going to spend some time working on the material you just received. You will have twelve minutes to spend on this material. Start working on the material you have now; we will stop you in twelve minutes. If you have any questions, raise your hand, and the experimenter will see you.

(After the twelve minutes)
We will now take a test on the material covered. Leave the test face down until I tell you to start. The test is in two parts. The first part is completion and short-answer questions; the second part is multiple-choice. You will not need the sponge and black plastic stylus in the first part, so lay the sponge and black stylus on the floor under your chair. You will have 55 minutes on this test. Try to answer every question. Do not spend too much time on any one question. If you are not sure of the answer, write down your best guess and continue to the next question. When you complete the top test, the short-answer test, raise it in your hand and the experimenter will come around and take it from you. As soon as he has taken the first test from you, turn over the multiple-choice test, read the instructions and proceed to answer the multiple-choice questions. To repeat, you have a total of 55 minutes for both tests. Do not spend too much time on one item; keep moving. As soon as you have completed the short-answer test, raise it in your hand. As soon as the experimenter has taken your first test, continue on to the multiple-choice test. When you are finished with both tests, you may leave. Do your best, but try to finish all items in both tests. If you have any questions, raise your hand after we start. Ready, turn the top test over and start.

Week 2--Groups C1 and C2.

Our procedure today will be somewhat different than last week. First you will work on the printed material you have; then we will see a film on a descriptive statistics topic.

APPENDIX C (continued)

Pay attention to the film and learn as much from it as you can. After the film is over, you will be tested on the material covered in the film. Before you see the film, however, we will spend some time working on the materials you just received. Start working on the material you have now; we will stop you in twelve minutes. If you have any questions, raise your hand, and the experimenter will see you.

(After the film and rest)

We will now take a test on the material covered in the film. Leave the test face down until I tell you to start. The test is in two parts. The first part is completion and short-answer questions; the second part is multiple-choice. You will not need the sponge and black plastic stylus in the first part, so lay the sponge and black stylus on the floor under your chair. You will have 55 minutes on this test. Try to answer every question. Do not spend too much time on any one question. If you are not sure of the answer, write down your best guess and continue to the next question. When you complete the top test, the short-answer test, raise it in your hand and the experimenter will come around and take it from you. As soon as he has taken the first test from you, turn over the multiple-choice test, read the instructions and proceed to answer the multiple-choice questions. To repeat, you have a total of 55 minutes for both tests. Do not spend too much time on one item; keep moving. As soon as you have completed the short-answer test, raise it in your hand. As soon as the experimenter has taken your first test, continue on to the multiple-choice test. When you are finished with both tests, you may leave. Do your best, but try to finish all items in both tests. If you have any questions, raise your hand after we start. Ready, turn the top test over and start.

Pre Test Freq. Dist.
F 63
Statistics Exp.

Name________________
Class________________
Today's Date__________

Record answers on test in the blanks provided.

1. X: 1, 3, -2 symbolizes

   a) an unknown, and three knowns.
   b) three unknowns.
   c) three sets of X data, with n = 1 for each.
   d) one set of X data, with n = 3.
   e) none of the above.

1.________

2. The best way of condensing 1000 test scores into more concise form is to

   a) throw away extreme values.
   b) list each person in rank order of his score.
   c) list each score value and the number of times it occurred.
   d) list the number of different score values that occurred.
   e) separate the very high scores from the low ones.

2.________

3. It is permissible to have some scores fall into two classes of a frequency distribution when the

   a) scores are grouped.
   b) number of observations is large.
   c) distribution is widely spread out.
   d) all of the above.
   e) none of the above.

3.________

4. A used-car dealer tests each of his 1960 model cars for gas- mileage. These values, when rounded to the nearest mile per gallon, ranged from 10 to 28. He compiled a proper frequency table of mileage and noticed that no car got a score of 25 mi/gal. He could tell because

   a) that category was empty, no frequency was recorded.
   b) that category was omitted.
   c) the frequency listed was zero.
   d) he remembered from testing -- the frequency table couldn't contain that information.
   e) none of the above.

4.________

5. A mathematical statistician, in paging through a book, finds this interesting random number: 2044347151192165312.
Make a frequency distribution of the digits in the number.

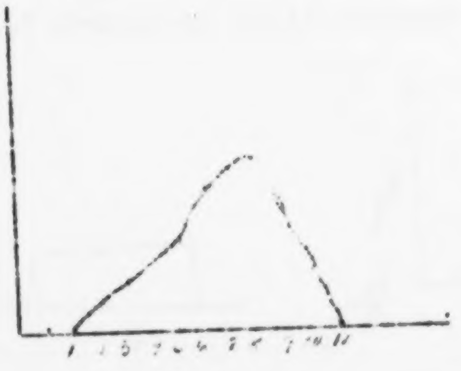
6. If he plotted this frequency distribution as a graph, he would place ____________ values on the horizontal axis and _________ values on the vertical axis.

7. Convert the following frequency distribution table into a frequency polygon (i.e., a graph of the frequency distribution).

X	f
10	5
9	8
8	12
7	9
6	7
5	6

Be sure to lable both axes, and fill in the numbers needed at the unit marks.

8. The curve below is best described as a

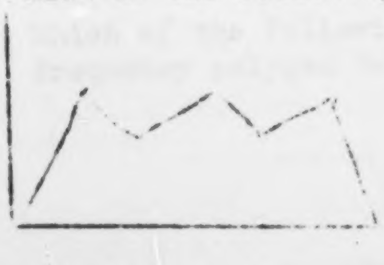


a) bi-skewed curve.
b) rectangular bell-shaped curve.
c) positively skewed curve.
d) bimodal curve.
e) frequency polygon.

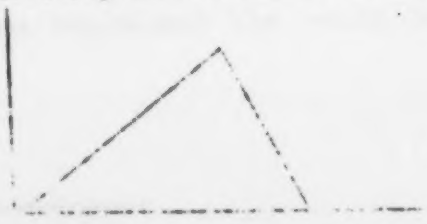
8.________

9. Which of the following is an example of a symmetric, multimodal distribution?

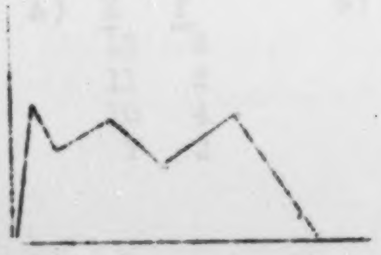
a)



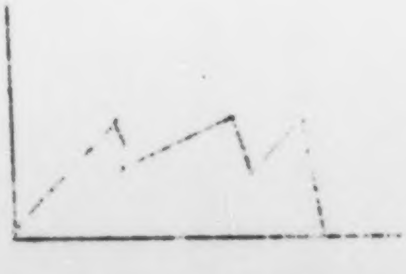
b.)



c)



d)



E e)none of the above.

9.________

10. When a frequency distribution has increasingly large numbers of cases falling at the extremes, it will form a

a) curve skewed to the right.
b) curve skewed to the left.
c) U-shaped curve.
d) rectangular curve.
e) bell-shaped curve.

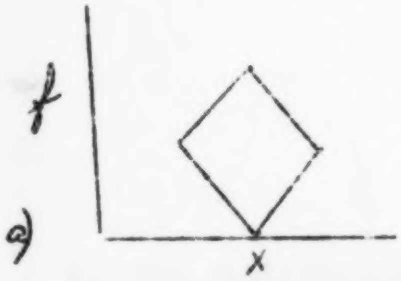
10.______

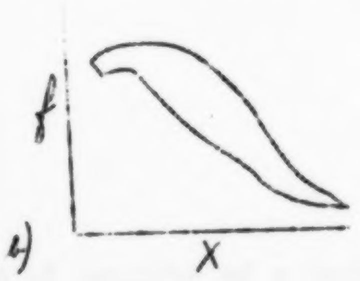
11. Mechanical comprehension test scores for adult men and women in the U.S. would form what kind of a distribution?

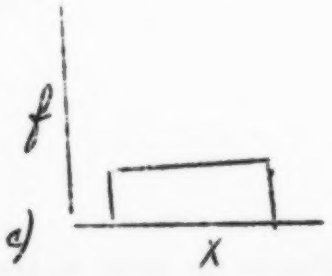
a) rectangular.
b) normal bell-shaped.
c) bimodal.
d) negatively skewed.
e) impossible to predict.

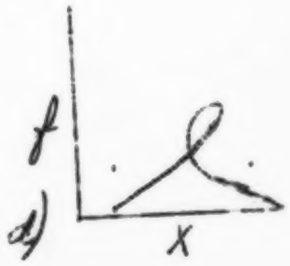
11.______

12. Which of the following would make a meaningful frequency distribution?







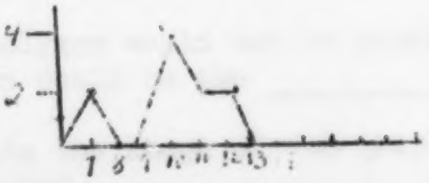


e) none of the above.

12.______

13. A bell-shaped distribution has one__________and two__________.

14. Which of the following frequency distributions represent the numerical frequency polygon below?



A)

X	f
12	2
11	2
10	4
7	2

b)

X	f
2	12
2	11
4	10
2	7

c)

X	f
12	2
11	2
10	4
9	0
8	0
7	2

d d) the frequency polygon is inaccurate because 2 and 4 are listed on the vertical axis.

e) a frequency polygon should not have been used on these data.

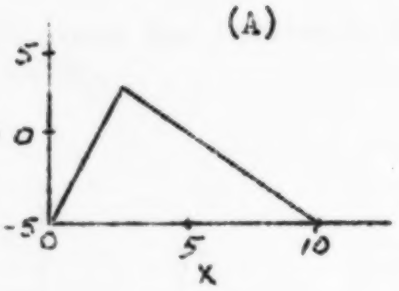
14.______

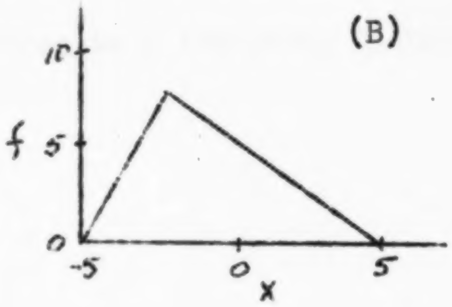
POST-TEST FREQUENCY DISTRIBUTIONS
OE F63
Statistics Exp.

DATE ________________

NAME ________________
GROUP ________________
COURSE ________________

1. Explain the difference between Figure A, and Figure B below, (in a few words).





2. The distribution of heights of a population made up equally of collegiate basketball players and jockeys (professional riders of race horses) would be: ________________.

3. A survey is conducted of all of the records of known polio patients in the U.S. from 1910 to the present. The age at which polio was contracted is recorded. The form of the curve of this data is (was) ________________.

4. A diamond-shaped frequency polygon would not be possible because, for a given ____________ there would be two ____________.

5. In the following table of data the shape of the graph would be described as ____________ and ____________.

The common name for this distribution is ____________.

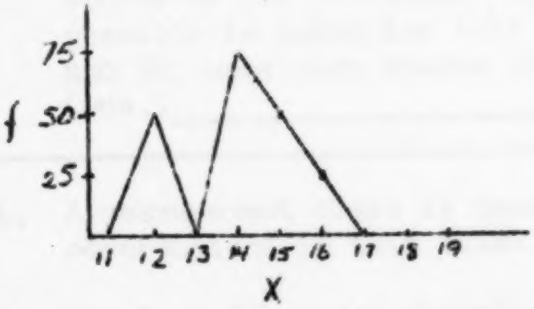
X	f
1	16
2	14
3	12
4	10
5	8
6	8
7	10
8	12
9	14
10	16

6. When a frequency distribution has increasingly large numbers of cases falling at the ______________ side, it will form a curve ______________ to the left.

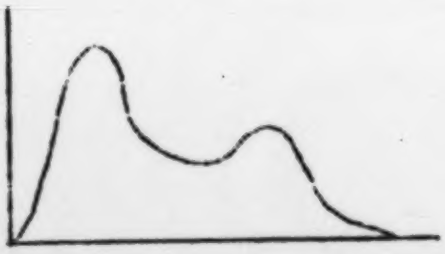
7. Identify the shape characteristics of the frequency distribution of the following scores: -4,-1,0,1,1,1,2,2,2,3,3,3,4,5,8. This distribution is _______modal and ______________.

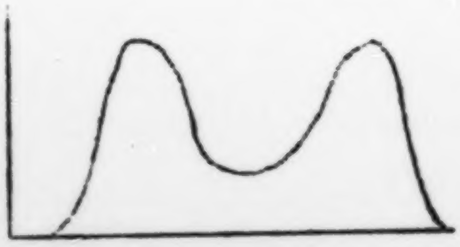
8. Describing a frequency polygon as ____-shaped, _____-shaped, or _____-shaped, indicates the location of the most frequent scores.

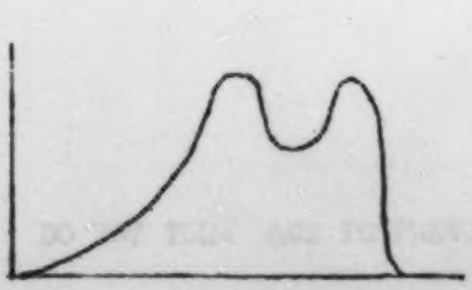
9. Convert the following frequency polygon to a frequency distribution table.

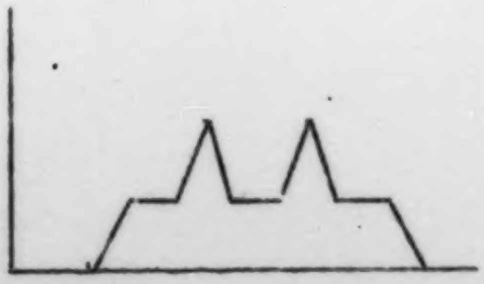


10. The entire collection of frequency distributions shown below can be labeled as _________________________.









DO NOT TURN BACK TO THE PREVIOUS PAGE

11. Given the following set of scores, construct a frequency ploygon (i.e., a graph describing the set of data). 3,3,6,4,4,6,7,7,7,7,8,8,9.

Be sure to label both axes, and fill in the numbers at the units marks.

12. In a frequency ploygon, ________ are placed on the horizontal axis and ___________ on the vertical axis.

13. A proper frequency distribution table was prepared by a teacher on the scores of his students. When the distribution was studied, it was possible to determine that while several students made scores of 95 and 92, none made scores of 93 or 94. This was clearly show by (wds.)____

____

14. A measurement class is usually symbolized as _____ and the number of observations in this class is symbolized as _________.

15. Why is a frequency distribution preferred over simply a list of the entire set of original scores?

DO NOT TURN BACK TO PREVIOUS PAGES

16. On a 50 point quiz, the following scores were obtained: 15, 21, 21, 25, 28, 30, 31, 32, 34, 35, 36, 40, 41, 42, 44 46, 50, 50. Describe everything that is wrong with the frequency table below (if anything).

X	f
35-49	7
30-39	6
20-29	4

The following scores are obtained on a 15 point test. These are used on all remaining questions.

John	10	Jack	9	Jim	7
Bill	8	Les	4	Rick	10
Mary	9	Avis	8	David	9
Fred	11	Art	3	Cleo	10
Joan	6	Dick	9	Pete	8
Rita	8	Paul	8	Bob	9

17. In the above set of data X= ______(wd.) and n = ______(#).

18. Construct a proper table for the numbers in the set of data.

19. If this table has been constructed properly, then the _____of the frequencies will _________________.

20. Construct a graph (frequency ploygon) which represents the set of data. Be sure to identify the axes, and the numerical values on each axis.

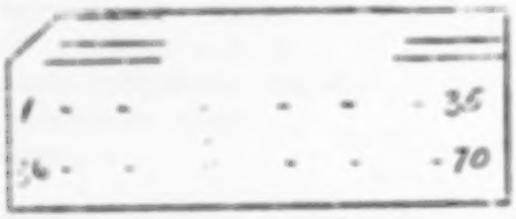
21. The shape of the graph you have constructed tells you that the form is ____________________ and also________________________.

DO NOT TURN BACK TO PREVIOUS PAGES..Turn this test in to the assistant before you go on to the multiple choice test.

Test Instructions.

The questions in this test are all multiple-choice. To record your choice you will need (1) your own pen or pencil, (2) an answer card, (3) a plastic stylus and (4) a sponge. (If you do not now have all these materials, ask the instructor to provide them).

Place the answer card in the following position on a smooth flat surface:



Print your last name and first name on the answer card. In the course space write "Statistics Experiment".

In the space after TEST write F.D. [redacted] Fill in today's date, in the form month, day, '63 as on the board. Under Misc. write the letters and numbers that appear on your reminder card after "Group".

There are 38 questions in this test. The numbers 1 to 35 at the top of the card represent the first 35 test items, and the letters A to E represent the answer choices of the questions. The last three items, 36 to 38 are found in the lower half of the card, on the left. Read each question and select your answer. Record your selection by circling the letter of the answer you think is correct beneath the number of the question. Be certain that you answer all 38 items, circling the correct responses on the answer card. DO NOT MARK OR WRITE ON THIS TEST BOOKLET. Use the scrap paper for any calculations, and indicate your answers by circling on the answer card. If you want to change an answer, DO NOT ERASE, merely make an "X" through your mistake and circle your new correct response.

After you have circled your answers for all 38 questions, take your answer card, and place it on the sponge pad provided. Using the stylus provided, carefully punch out ALL circled responses.

DO NOT PUNCH ANY CIRCLE WITH AN "X" THROUGH IT. Be sure the punched-out portion is detached. When you finish, turn your card over, and clean off any remaining loose portions of punched answers.

Be careful. Check the question number in the test with the number on the answer card each time you circle an answer. It is easy to lose your place. Should you make an irrepairable mistake, raise your hand, and the assistant will give you a new card.

POST-TEST FREQUENCY DISTRIBUTIONS
MC F63
Statistics Exp.

Do Not Record Answers on Test.

Circle answers on the IBM card. Use scrap paper for computations.

1. A person has a frequency distribution when he has obtained

   a) a record of every score value.
   b) the measurement classes and n.
   c) n scores for each frequency class.
   d) f for each measurement class.
   e) none of the above.

2. You want to compare the size of several sets of data consisting of 50 numbers each to see which is the largest. The best way would be to

   a) find the highest and lowest scores in each.
   b) put all of the scores of each in order.
   c) compare the largest numbers in each.
   d) construct a frequency polygon for each.
   e) use pie charts for each.

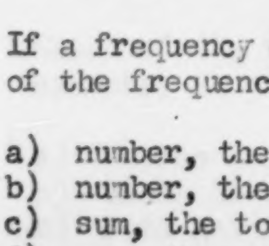
3. If a frequency distribution is properly set up, then the ________ of the frequencies must equal ________________________.

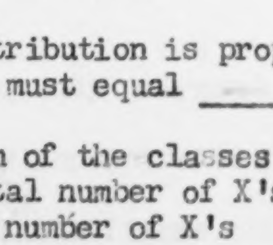
   a) number, the sum of the classes
   b) number, the total number of X's
   c) sum, the total number of X's
   d) sum, the number of classes
   e) none of the above

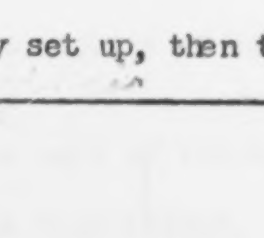
4. In a frequency distribution, each observation from the original set of data must

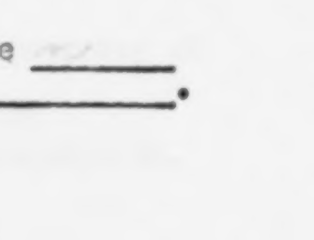
   a) be condensed into a more concise form.
   b) set down a record of the number of scores that fall in each score class.
   c) record the frequency of each measurement class.
   d) be included in one class and excluded from all other classes.
   e) take into account measurement and frequencies of each class.

5. Which of the following pairs of axes could be used to make a frequency polygon of scores on test V?

   a)


   b)


   c)


   d)


   e) all of the above or none of the above.

6. In a frequency distribution of 100 scores, f may

a) be as small as 0, and increase in units of .01.
b) be as small as 0, and increase in units of 1.
c) be as small as -100 and increase in units of .01.
d) be as small as -100 and increase in units of 1.
e) none of the above.

7. Select the proper frequency distribution for the following set of data: 14, 14, 16, 12, 15, 14, 12, 15, 15, 15

(a)

X	f
16	1
15	4
14	3
12	2

(b)

X	f
0	13
1	16
2	12
3	14
4	15

(c)

X	f
15	4
14	3
12	2
16	1

(d)

X	f
16	1
15	4
14	3
13	0
12	2

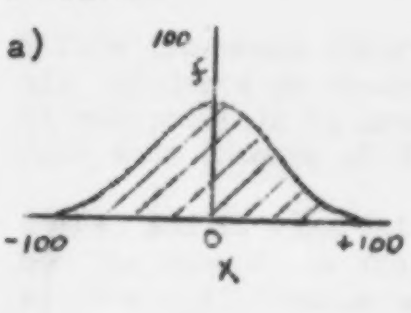
e) none of the above

8. Which of the following is the best description of the property of symmetry in a frequency polygon?

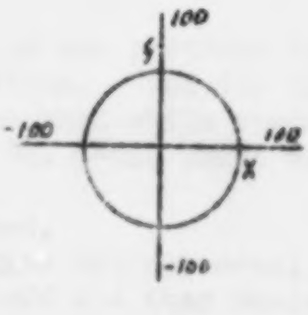
a) scores heavily concentrated on one part of the distribution.
b) looks like a U.
c) both halves of the figure coincide when folded.
d) looks like a bell.
e) equally long tails at each end.

9. Which of the following graphs are possible meaningful frequency polygons?

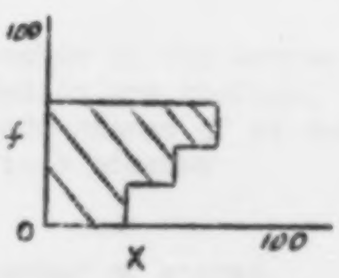
a)



b)



c)



d) all of the above
e) none of the above

10. To make a rectangular distribution multimodal, a person could

a) group every 2 X values so each interval would have twice the original frequency.
b) add a constant frequency to all X values.
c) add a constant frequency to every other X value.
d) add a frequency value to just the center X value.
e) add a frequency value to one extreme X value.

11. When plotting a frequency polygon, the point for each score class is plotted at or above the

a) upper limit of the score class interval.
b) mid-point of the score class interval.
c) lower limit of the score class interval.
d) average of the X values obtained in the score class interval.
e) none of the above.

12. Symmetrical distributions can't be

a) unimodal.
b) bimodal.
c) skewed.
d) smoothed.
e) flat.

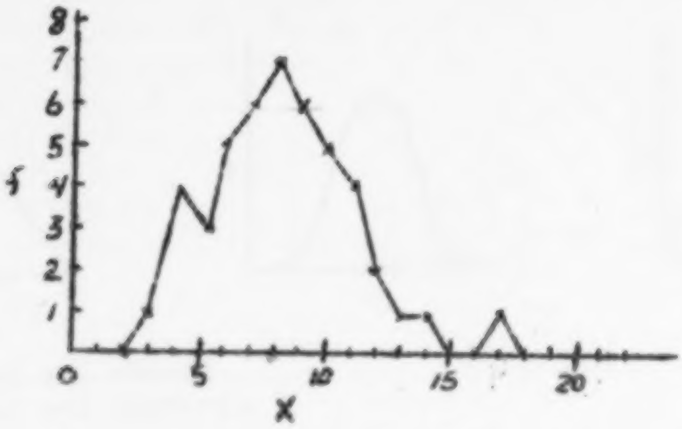
13. For a set of data X: -22, 5, 14, 0, -6, 5

a) n = 5 since n = any X value other than zero.
b) n = 5 since n = only the last X values.
c) n = 5 since n = the number of scores other than zero.
d) n = 5 since n = the most frequently occuring number.
e) none of the above.

14. A fine frequency distribution was prepared by a teacher on the scores of his students on an examination. When the distribution was studied, it was possible to determine that while students made scores of 95 and 92, none made a score of 93 or 94. This point was obvious because

a) 93 and 94 were not listed.
b) the sum of the frequencies did not equal the number of scores.
c) the distribution broke off and then resumed.
d) the classes of observation were unequal.
e) the frequency of the scores 93 and 94 was zero.

Use the frequency distribution shown to answer the following questions.



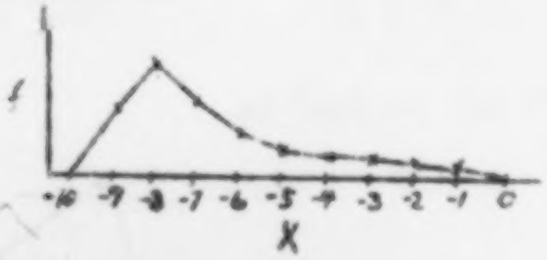
15. What is the highest score in the set of data? _____ the lowest score? _____.

a) 20,0
b) 18,1
c) 17,2
d) 8,0
e) none of the above pairs.

16. How many times did a score of 6 appear in the set of data?

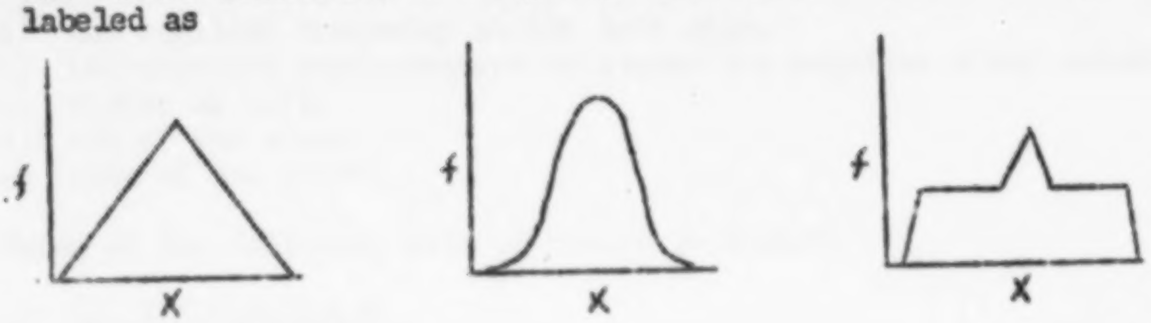
a) 6
b) 5
c) 3
d) 1
e) none of the above

17. Which of the following is an accurate description of the figure below?



a) Unimodal, positively skewed
b) Unimodal, negatively skewed
c) Multimodal
d) Bell shaped
e) None of the above
f) A nonsense figure, not a possible frequency polygon (punch out all 5 above).

18. The entire collection of frequency distributions shown below can be labeled as

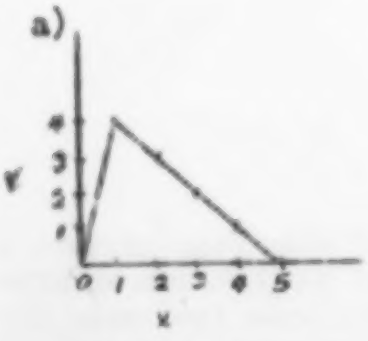


a) unimodal and skewed
b) unimodal and symetric
c) multiskewed
d) multi-symetric
e) none of the above

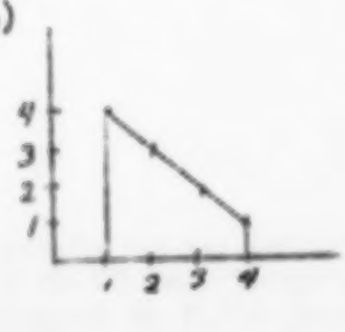
19. Which of the frequency polygons below correctly represents the following data?

X	f
4	1
3	2
2	3
1	4

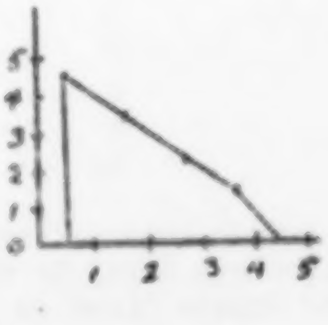
a)



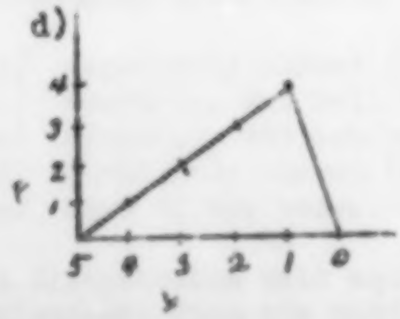
b)



c)



d)



e) none of the above is correct

20. The distribution of the distances that a group made up equally of 12-year-old boys and professional baseball players can throw a baseball would be

a) bell shaped
b) bimodal
c) negatively skewed
d) positively skewed
e) unpredictable

21. A positively skewed distribution may have

a) the smallest frequency at the left side.
b) the greatest concentration of scores for negative score values.
c) scores of zero.
d) all of the above.
e) none of the above.

22. Which of the following sets of scores is bimodal?

a) 3,4,5,5,6,6,7,8,12
b) 3,3,3,6,9,12,3,3,3
c) 3,3,3,6,7,9,12,12,12
d) all of the above.
e) none of the above

23. The following frequency table shows the minimum speed recorded on various cars as they drove past a certain stop sign. A frequency polygon of this data would best be described as

a) symetrical, unimodal.
b) positively skewed, unimodal, J shaped.
c) negatively skewed, unimodal, J shaped.
d) meaningless.
e) none of the above combinations are correct.

Min. Speed (mph)	f
0	50
1	15
2	9
3	7
4	5
5	2
6	2
7	0
8	1
9	0
10	1

24. A reading test with 50 possible points, yields a bell-shaped distribution with scores ranging from 5 to 48 on a large sample of third graders. If the same test were administered to fifth graders, we would expect the frequency distribution to be:

a) negatively skewed (skewed to the left).
b) symetric, and bell shaped.
c) symetric, but not bell shaped.
d) positively skewed (skewed to the right).
e) none of the above.

25. A distribution with equal frequencies at all points that are equal distances from the middle class could be

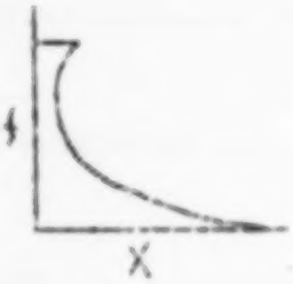
a) symmetric.
b) bimodal.
c) rectangular.
d) bell-shaped.
e) could be any one of the above.

26. The frequency polygon for incomes of U.S. taxpayers would be

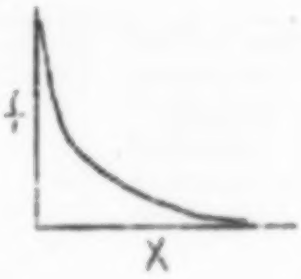
a) skewed to the right (positively) during a depression.
b) skewed to the left (negatively) during a depression.
c) symmetric and multimodal.
d) J-shaped.
e) bell-shaped.

27. Which of the following is an example of a negatively skewed, unimodal, J-shaped distribution?

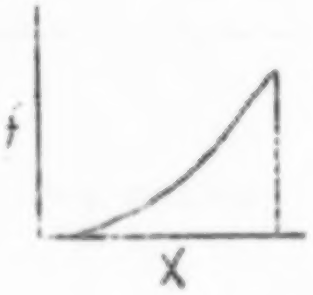
(a)



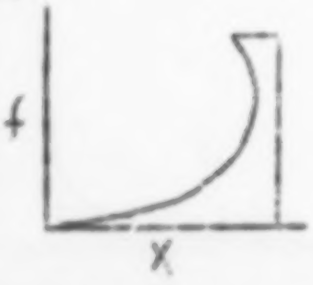
(b)



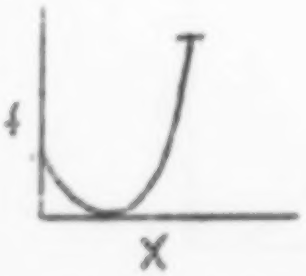
(c)



(d)



(e)



28. A gasoline station attendant records the dollar-value of cash gas sales for several days. The distribution table was:

Price	f
.01 - .50	2
.51 - 1.00	30
1.01 - 1.50	50
1.51 - 2.00	20
2.01 - 2.50	30
2.51 - 3.00	60
3.01 - 3.50	85
3.51 - 4.00	12
4.01 - 4.50	1

The form of the frequency polygon is

a) unimodal skewed positive.
b) multimodal skewed positive.
c) unimodal skewed negative.
d) multimodal skewed negative.
e) none of the above.

29. The following distribution would best be described as X: -5,-3,-2,-2, -1,-1,-1,-1,0,0,1,3

a) symmetric, unimodal
b) negatively skewed, unimodal
c) positively skewed, unimodal
d) impossible.
e) none of the above.

. Items 30 through 35 have 10 possible responses. On these items, we will use the entire column, from the top (starting at the item number) all the way to the bottom (ignoring the number in the bottom half). Each graph below is a question to which you are to mark all correct responses. i.e., for each frequency polygon below, mark all appropriate responses. You may, therefore, mark more than one response for each column.

RESPONSES

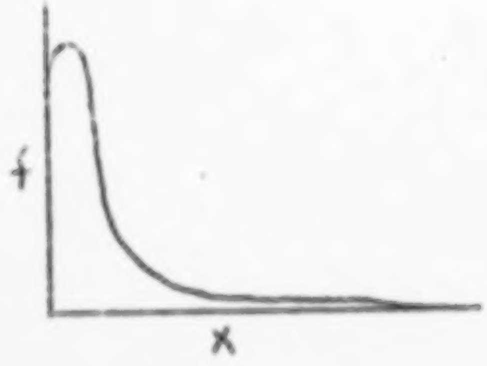
UPPER section of card

a) positively skewed (skewed to the right)
b) symmetric
c) negatively skewed (skewed to the left)
d) unimodal
e) multimodal

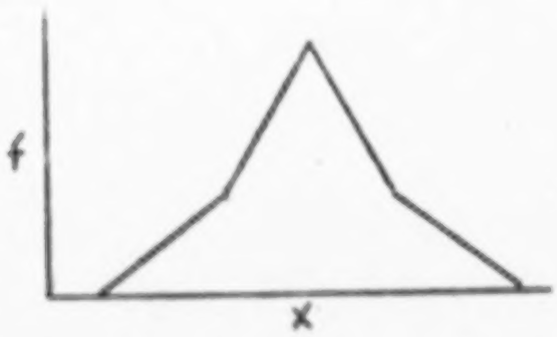
LOWER section of card

a) bell-shaped
b) J-shaped
c) U-shaped
d) S-shaped
e) a nonsense figure, not a possible frequency polygon.

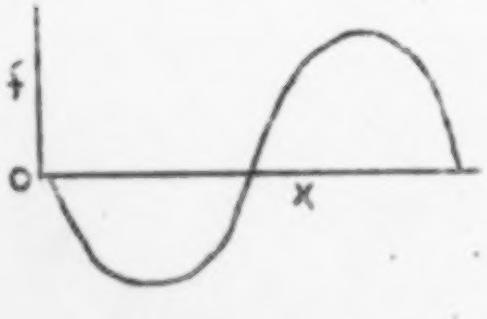
30.



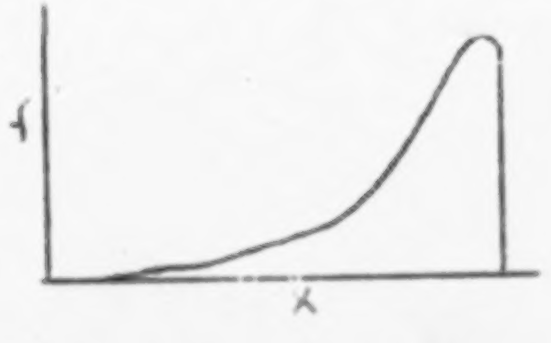
31.



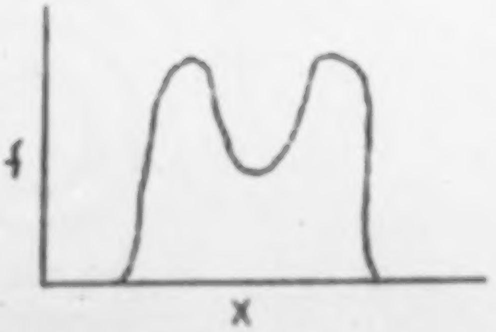
32.



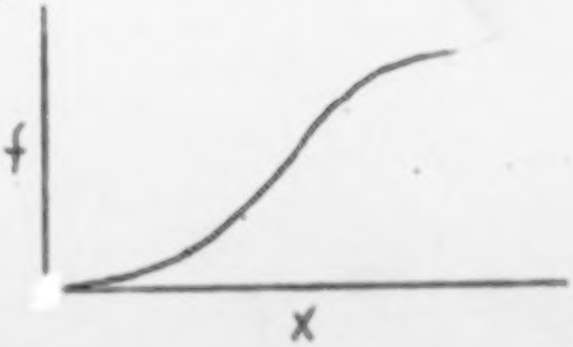
33.



34.



35.



36. A survey is conducted about the bed habits of U.S. citizens, in which information is gathered about 100 persons of each yearly age group from 0-1 to 89-90. For each group, we record the number of the 100 persons who spend over $\frac{1}{2}$ of each day in bed. When plotted as a frequency distribution, with age as the baseline, what will be the shape of this distribution?

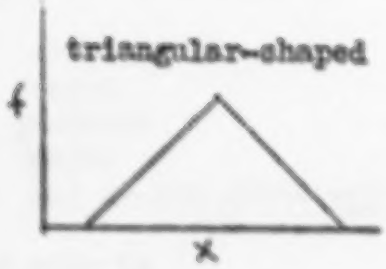
a) skewed to the right (positively)
b) skewed to the left (negatively)
c) U-shaped
d) bell shaped
e) none of the above

37. A distribution with the largest frequency in one and only one class may be

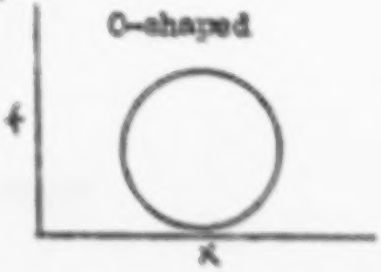
a) skewed.
b) J-shaped.
c) unimodal.
d) symmetric.
e) any of the above.

38. Which of the following is a meaningful representation of a frequency distribution?

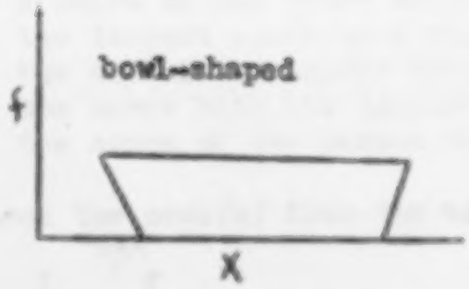
a)



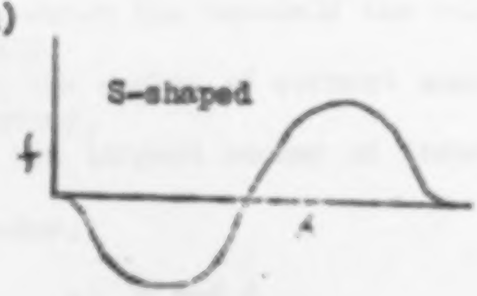
b)



c)



d)



e) none of the above or two or more of the above

Pre Test -- Central Tendency
MC F63
Statistics Exp.

Name____________________
Class____________________
Today's Date____________________

record answers on test.

in blanks provided.

1. A central tendency index is

   a) a point on a score scale.
   b) a distance on a score scale.
   c) an area above a score scale.
   d) a point on the frequency curve.
   e) a point on the frequency scale.

____ 2. Central tendency indices describe the

   a) general location of a set of data.
   b) tendency to high or low scores in a set of data.
   c) tendency to average scores in a set of data.
   d) form of a distribution of data.
   e) none of the above.

____ 3. Which of the following are not specific indices?

   a) averages
   b) means
   c) medians
   d) modes
   e) all of the above are specific indices

____ 4. The mode is

   a) a point on the score scale below which lie one-half the scores.
   b) the largest score in a set of data.
   c) the sum of the scores divided by the number of correct answers.
   d) the score with the largest frequency.
   e) the score of the person who got the largest number of items correct.

____ 5. Select the mode(s) from the table below.

X	f
31	2
30	2
29	6
28	4
27	6
26	3
25	1

   a) 6 and 6
   b) 29
   c) 28
   d) 27 and 29
   e) none of the above

____

6. The median is the

a) midpoint between the largest score and the smallest score.
b) midpoint of the score class that contains 50% of the cases.
c) score point that has as many cases below it as above it.
d) average of the larger and smaller scores.
e) none of the above.

______

7. Select the median of the following set of data: 4,5,6,6,6,6,7,7,7,8,8,9,9,10.

a) 6.5
b) 7.0
c) 7.2
d) 7.5
e) none of the above (within ± .01)

______

8. The mean is a

a) percent.
b) score value.
c) distribution.
d) hypothetical number.
e) central limit.

______

9. Compute the mean of the following scores on a 17 item test scored +1 for right, 0 for blanks, and -1 for wrong answers.
12, 8, 8, 7, 7, 3, 0, -5, -10, -13

a) 1.0
b) 1.7
c) 2.1
d) 8.1
e) none of the above (within ± .1)

______

10. A measure of central tendency which depends on every score value in the distribution is the

a) mean.
b) median.
c) mode.
d) all of the above.
e) none of the above.

______

11. The mean should be used as the measure of central tendency for a distribution whenever a point reflecting the

a) most 'typical' score is desired.
b) most 'frequent' score is desired.
c) overall numerical total is desired.
d) overall form of the distribution is desired.
e) none of the above.

______

12. The median is the preferred measure of central tendency when

a) the distribution is symmetrical.
b) some score values are absent.
c) the values of all scores are to be taken into account.
d) a multivalued index is desired.
e) there are no extreme values involved.

13. The mode should be used in preference to the median and the mean

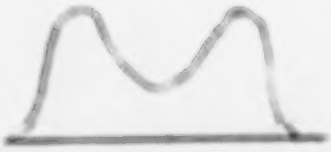
a) only when a multivalued index is usable.
b) only in a multimodal distribution.
c) only when the modal frequency is clearly larger than other frequencies in its neighborhood.
d) all of these are necessary.
e) none of these is necessary.

14. When a score in a distribution is changed, the ________________ will always change.

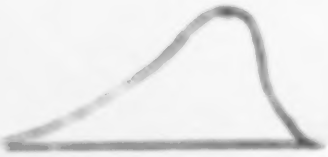
a) median
b) mean
c) mode and median
d) mode, median and mean
e) none of the above

15. In which of the distributions below are the mean and median exactly the same score?

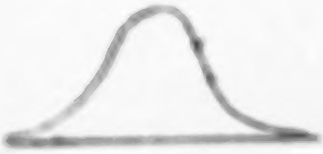
(a)



(b)

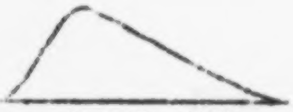


(c)



d) both a and c

e) no such distribution exists.

16. In a distribution skewed to the right,  the three measures of central tendency yield scores of 101, 103, and 105. Which measure has which score?

a) mode = 101, mean = 103, median = 105
b) mean = 101, median = 103, mode = 105
c) mode = 101, median = 103, mean = 105
d) median = 101, mode = 103, mean = 105
e) mean = 101, mode = 103, median = 105

____
17. Finding the largest frequency distribution is the first step for finding the __________ as a measure of central tendency.

Post Test -- Central Tendency
OE F63
Statistics Exp.

Date ____________

Name ____________________
Group ____________________
Course ____________________

1. The specific indices of central tendency are the ________________, the ________________, and the ________________.

2. A central tendency index is best thought of as a single ____________ on the score scale to represent the ________________ of a complete set of data.

3. The mode is the ____________ with the largest ____________.

4. The median is defined as ____

____

____.

5. The mean is defined as ____

____

____.

6. The mode for the table below is ________.

X	f
10	5
9	7
8	8
7	9
6	10

7. The ____________, a measure of central tendency, is based on the numerical value of every ____________.

8. Compute the median and mean of the following set of scores.
-3, 0, 0, 0, 1, 3, 4, 4, 5, 8, 15.

Mean = ________ Median = ________

9. In measuring the heights of all the first graders in a certain class, the experimenter also measures the teacher. This would raise the ____________, as index of central tendency.

10. When a frequency distribution is exactly symmetrical, we know that ____

____.

11. A bowling score which occurs most often in the course of a season of bowling would be the bowler's ______________.

12. The Army tests the length of wear of two fabrics. Compiling the scores of many samples of each fabric, fabric A has the higher mean (longer mean wear) while fabric B has the higher median. Assuming they cost the same, and are equal on other factors, which brand should the Army buy? Why?

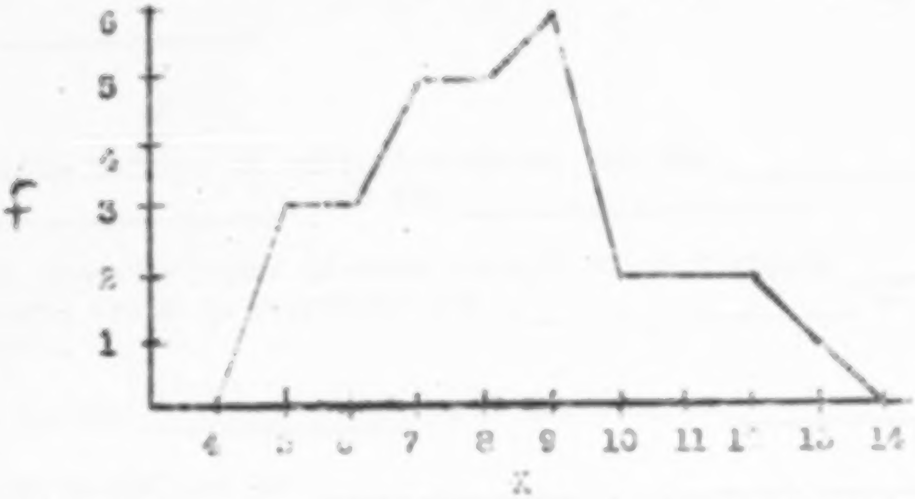
13. The annual income of all families of a town is reported, and the per capita income is reported as 8,000. The median income is only 4,000. Explain how this is possible.

14. In a situation like this, which measure of central tendency best represents the 'typical income'?

15. In grading Master's comprehensive examinations, one department uses a score system from 1 to 9. Each person reading the exam gives it one of the above numbers, then the mean of the scores is used to determine if the student passed or failed. One "expert" comments that using the mean in the above situation is not proper, since a reader who marks the paper a 9 has nine times the effect on the sum as does a reader who marks the paper with a 1. Consequently, he says, the person marking the paper a 9 has nine times more effect on the mean than does the person marking a paper a 1, and therefore he has a much larger role in determining whether the student passes or fails than does the person who marks tha paper a 1.

Do you agree with this "expert," that marking the paper with a 9 has a greater influence on the mean than does the person marking it a 1? Yes or No? Explain.

16. List modal value(s) for the following distribution: ______________.



17. Under what circumstances will the mode be the preferred index of central tendency?

Post Test -- Central Tendency
OE F63
Statistics Exp.

Name ____________________
Group ____________________
Course ____________________

Date ______________

1. The specific indices of central tendency are the ______________, the ______________, and the ______________.

2. A central tendency index is best thought of as a single ______________ on the score scale to represent the ______________ of a complete set of data.

3. The mode is the ______________ with the largest ______________.

4. The median is defined as ____
____
____.

5. The mean is defined as ____
____
____.

6. The mode for the table below is ________.

X	f
10	5
9	7
8	8
7	9
6	10

7. The ______________, a measure of central tendency, is based on the numerical value of every ______________.

8. Compute the median and mean of the following set of scores.
-3, 0, 0, 0, 1, 3, 4, 4, 5, 8, 15.

Mean = __________ Median = __________

9. In measuring the heights of all the first graders in a certain class, the experimenter also measures the teacher. This would raise the ______________, as index of central tendency.

10. When a frequency distribution is exactly symmetrical, we know that ____
____.

11. A bowling score which occurs most often in the course of a season of bowling would be the bowler's ______________.

12. The Army tests the length of wear of two fabrics. Compiling the scores of many samples of each fabric, fabric A has the higher mean (longer mean wear) while fabric B has the higher median. Assuming they cost the same, and are equal on other factors, which brand should the Army buy? Why?

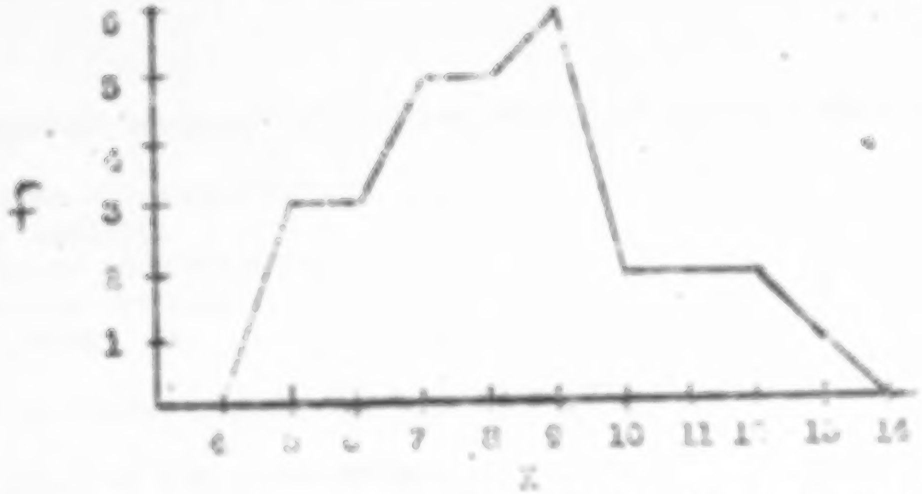
13. The annual income of all families of a town is reported, and the per capita income is reported as 8,000. The median income is only 4,000. Explain how this is possible.

14. In a situation like this, which measure of central tendency best represents the 'typical income'?

15. In grading Master's comprehensive examinations, one department uses a score system from 1 to 9. Each person reading the exam gives it one of the above numbers, then the mean of the scores is used to determine if the student passed or failed. One "expert" comments that using the mean in the above situation is not proper, since a reader who marks the paper a 9 has nine times the effect on the sum as does a reader who marks the paper with a 1. Consequently, he says, the person marking the paper a 9 has nine times more effect on the mean than does the person marking a paper a 1, and therefore he has a much larger role in determining whether the student passes or fails than does the person who marks tha paper a 1.

Do you agree with this "expert," that marking the paper with a 9 has a greater influence on the mean than does the person marking it a 1? Yes or No? Explain.

16. List modal value(s) for the following distribution: ______________.



17. Under what circumstances will the mode be the preferred index of central tendency?

Post Test -- Central Tendency
NC F63
Statistics Exp.

Do not record answers on test.
Circle answers on the IBM card.
Use scrap paper for computations.

1. The "central tendency" of a distribution of scores refers to its

   a) degree of concentration.
   b) variability.
   c) average relationship.
   d) average frequency.
   e) location.

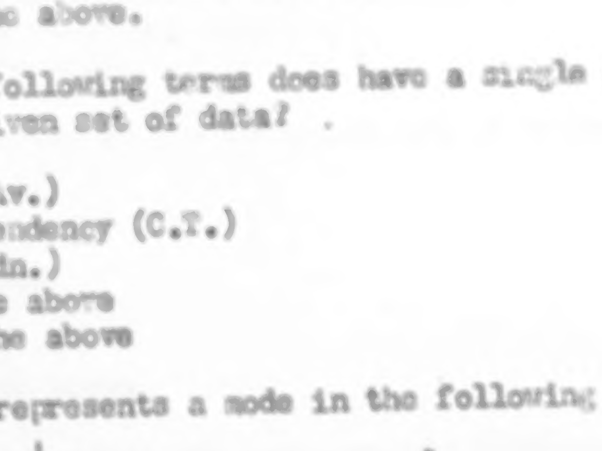
2. A central tendency index is

   a) a point on the score scale.
   b) a point on the frequency curve.
   c) a point on the frequency scale.
   d) an area above the score scale.
   e) none of the above.

3. Which of the following terms does have a single specific numerical value for a given set of data?

   a) average (Av.)
   b) central tendency (C.T.)
   c) median (Mdn.)
   d) all of the above
   e) none of the above

4. Which number represents a mode in the following frequency distribution?



   a) 5
   b) 8
   c) 10
   d) 16
   e) none of the above

5. A measure of central tendency which depends only on the number of smaller and larger scores (i.e. smaller and larger than the index itself) is the

a) mean.
b) median.
c) mode.
d) average.
e) all of the above.

6. Compute the mean of the following set of data: +1, 3, 0, 3, [illegible], -3, -3.

a) 0.0
b) -0.5
c) -1.0
d) -3.0
e) none of the above (within ± .1)

7. A single central tendency index is frequently computed because such an index is useful in

a) giving an indication of the number of classes.
b) giving an indication of the form of the distribution of frequencies.
c) giving an indication of the location of an entire set of data.
d) all of the above are true.
e) b and c are true.

8. Compute the median of the following set of data: -2, 3, 4, 4, -2, 2, -3, 3, -4, 1, -4, 4, -3, 3.

a) 3.0
b) 3.0 and -3.0
c) 0.0
d) -.5 and .5
e) none of the above

9. If the mean and median are considerably different you know that:

a) the distribution is not symmetrical
b) the mode lies between them
c) there are two different populations involved
d) the statistician made a mistake
e) none of the above

10. If a score in a distribution is increased by two points, this change will have the greatest effect on the mean if this score was near the

a) extremes of the distribution.
b) middle of the distribution.
c) original mean.
d) original mode.
e) all of above are equal in effect.

11. The score value at the most frequently occuring score in a distribution may be the _________________, but must be a __________________.

    a) median; center score value
    b) median; mean
    c) mode and mean; median
    d) mode; mean and median
    e) mean and median; mode

12. The Chamber of Commerce in a town that has many small house owners and 6 large industrial plants wants to attract new residents. In reporting the "average property tax income available for support of schools, roads, and services" to prospective parents, they would best use the

    a) mean.
    b) median.
    c) mode.
    d) b or c but not a.
    e) a or c but not b.

13. In the same town, when reporting the "average property tax paid" to a prospective house builder who wants to keep expenses down, they would make the most favorable impression by using the

    a) mean.
    b) median.
    c) mode.
    d) a or c but not b.
    e) b or c but not a.

14. If several authentic mathematical wizards are included in a college mathematics class, their scores would affect which measure of central tendency the most?

    a) mean
    b) median
    c) mode
    d) depends on the number of wizards
    e) depends on the number in the class

15. The operator of a fleet of cars finds he saves money by driving on his tires until all tread is gone. He tests the tread life of many samples of tire A, tire B and tire C. He finds the following confusing results.

	Highest	Middle	Lowest mileage
modes	A	B	C
medians	B	C	A
means	C	A	B

To save the most money, he would:

a) buy only tire A
b) buy only tire B
c) buy only tire C
d) buy A for city traffic and C for travel
e) the above is impossible

16. The median of a set of scores is defined as the score value that

a) is ½ the distance between the two most extreme scores.
b) is in the middle of the middle score category.
c) is at the point of balance.
d) has as many cases to it's left as it has to it's right in the frequency distribution.
e) has the sum of the deviations to it's left equal to the sum of the deviations to it's right.

17. The measure of central tendency that will always be changed if one of the scores in a distribution is changed is

a) the center value.
b) the center frequency or median value.
c) the modal value.
d) the mean value.
e) none of the above.

18. In a distribution of 200 heights of adults, measured to the nearest ½ inch, the mean and median are not exactly the same, but differ only by .03 inches.

The form of the distribution is probably

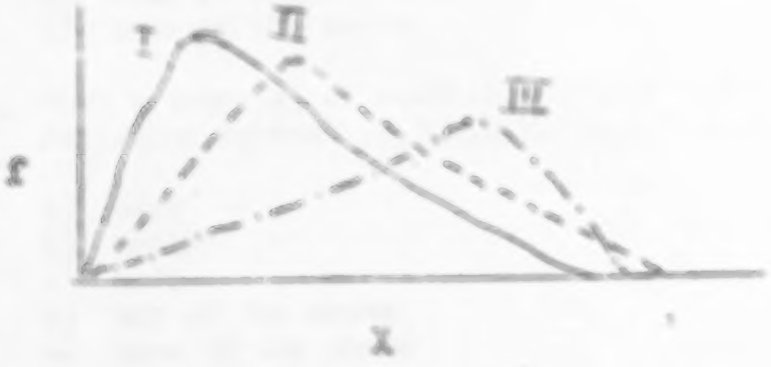
a) symmetric.
b) bimodal skewed.
c) unimodal skewed.
d) U-shaped.
e) impossible to say.

19. An experimenter desired to determine a statistical average of the time it took fifty rats to run a maze. 5 rats never finished the maze so he had no times on them. The experimenter

   a) should use the mean times of the 45 rats and assume that the missing data would have yielded the same approximate mean.
   b) should use the most frequently appearing time as the best statistical average.
   c) should use the median as the best statistical average.
   d) should average the mean and median as the best statistical average.
   e) it is impossible to determine a statistical average under these conditions.

20. Which distribution below has the largest numerical value of the mode?

   a) I
   b) II
   c) III
   d) all are equal
   e) impossible to tell from this data.

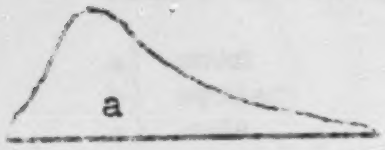


21. The following distribution has a mean of 16, and a median of 16. If scores of 15 and 30 are added to this distribution, what will happen?

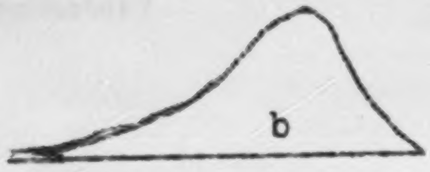
   12, 13, 14, 15, 15, 16, 16, 16, 16, 16, 17, 18, 20, 25, 16

   a) the mean, median, and mode will change.
   b) the mean, and median will change; the mode will not.
   c) the mean will change; the median and mode will not.
   d) the mean, median, and mode will not change.
   e) none of the above.

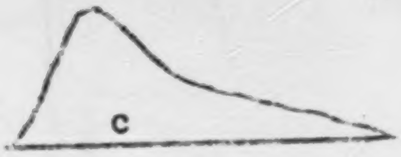
22. Reading from left to right below, in which of the following distributions are the median, mean, and mode properly ordered?



mode, mean, median



median, mode, mean



mode, median, mean

[illegible]
[illegible]
d) a and c
e) none of the above

23. The mean should be used to obtain a 'typical value' in a distribution

a) that is approximately symmetrical.
b) where several scores are missing.
c) that is clearly skewed.
d) that is clearly bimodal.
e) none of the above.

24. What measure(s) of central tendency would be most descriptive of the following scores: 2,3,4,4,4,4,4,6,8,9,11,12,12,12,12,12,13,15,15,17

a) mean
b) median
c) mode
d) any of the above
e) none of the above

25. An instructor computes the mean, median and mode for his first quiz of the term which yields a bell shaped distribution. An exceptionally bright student and an exceptionally dull student enter this large class of average students late and make up the quiz. This may change

a) the mean.
b) the mode.
c) the median.
d) all of the above.
e) b and c but not a.

26. In a symmetrical bimodal distribution

a) the modal, median and mean values are equal.
b) the median and mean values are equal.
c) the modal and median values are equal.
d) the mean and modal values are quall.
e) the modal, median and mean values are all different in value.

27. Which of the following would usually be used to indicate the typical salary of college graduates?

a) mean
b) median
c) mode
d) either a or c
e) either b or c

28. Suppose that there are 40,000,000 persons who have yearly income taxes worth considering. In this positively skewed distribution the statistician is concerned as to what index of central tendency he should use to reflect the tax effect upon the total government income. His decision should be:

a) mean
b) median
c) mode
d) b and c
e) none of the above

29. The most useful index of central tendency is the mode when

a) the distribution is markedly skewed.
b) a typical value is desired.
c) the distribution is symmetrical.
d) the numerical value of all scores is involved.
e) a multimodal situation occurs and multiple values are usable.

The following data was collected by the state highway patrol in a spot safety check of all cars on a state highway.

I	II	III	IV	V	VI
# of safety deficiencies.	Year of Car	Mileage of Car (in 1000's)	Drivers Occupation	Drivers General Income	Drivers Age
7	'55	83,	Laborer	3,200	36
2	'60	79,	Salesman	5,100	35
0	'60	52,	Teacher	5,300	36
0	'55	70,	Student	1,200	20
1	'63	38,	Salesman	4,500	20
0	'64	3,	Lawyer	15,400	20
2	'57	60,	Housewife	------	21
0	'56	72,	Clerk	5,200	37
3	'62	82,	Salesman	5,600	36
1	'61	65,	Student	3,000	20
0	'64	12,	M.D.	45,000	37
0	'58	78,	Minister	4,800	36
9	'56	71,	Housewife	1,000	22
0	'63	41,	Executive	9,000	38
2	'59	72,	Student	500	20
1	'59	76,	Laborer	4,100	20
0	'61	63,	Carpenter	6,100	23
1	'57	10,	Barber	4,600	19
5	'58	63,	Housewife	-----	22
0	'62	21,	Student	-----	19

30. Compute the mean of the number of safety deficiencies.
a) 0.0 b) 1.0 c) 1.7 d) 3.1 e) none of the above (within ± .1)

31. Compute the median drivers age.
a) 22.0 b) 22.5 c) 23.0 d) 23.5 e) none of the above (within ±.01)

32. On which measure must you use the median rather than the mean as a measure of central tendency?
a) II b) III c) V d) VI e) none of the above

33. On which measure would the mean be preferred to reflect the total outcome of all data obtained in the measure?
a) I b) III c) VI d) all of the above e) none of the above

34. On which measure would you prefer the mode as the most descriptive of the three indices?
a) I b) IV c) V d) VI e) none of the above

35. On which measure would the mean be [illegible] above the median?
a) I b) II c) III d) IV e) none of the above

36. If you were interested in reporting "typical" results, on which measure would you prefer the median to the mean?
a) I b) II c) VI d) all of the above e) none of the above

37. Which distribution(s) would be classified as multimodal?
a) I b) II c) III d) VI e) none of the above

38. On which measure would the mean be substantially lower than the median?
a) I b) II c) III d) V e) none of the above

39. On which measure would it be impossible to obtain a median?
a) I b) II c) IV d) V e) none of the above

40. Which distribution would be classified as J-shaped?
a) I b) II c) III d) VI e) none of the above

41. Which distribution would be classified as bell-shaped -- symmetrical?
a) I b) II c) III d) VI e) none of the above

APPENDIX D1

Frequency Distribution Immediate Post-Test--Open-End Scores

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (S)	1	958.62	14.970*
Presentation (P)	2	196.91	3.075*
Introduction (I)	2	327.58	5.115*
S x P	2	151.30	2.363
S x I	2	90.29	1.410
P x I	4	43.38	<1
S x P x I	4	60.99	<1
Within cells	316[1]	64.038	

1) To achieve proportional cell frequencies without dropping even more subjects than was done, a value equal to the cell mean was added to cells 111 and 133 (SPI factor values). These values would not contribute to the within cells sum of squares, and so were not counted in the degrees of freedom, reducing this from 318 to 316. (See Lindquist, 1953, p. 148) This same value is used in all of the Freq. Dist. Immediate Post-test analyses.

Cell means(first row), variances(second row), and n's (bottom row):

A Group

	P1	P2	P3
I1	18.08 70.25 24	16.83 59.10 24	15.67 55.80 24
I2	19.03 47.48 30	15.30 83.60 30	15.97 66.79 30
I3	15.80 60.79 30	17.07 67.03 30	11.40 35.77 30

B Group

	P1	P2	P3
I1	24.62 49.12 8	25.75 61.07 8	20.38 81.12 8
I2	16.90 74.10 10	21.90 108.10 10	15.30 106.90 10
I3	18.80 71.96 10	20.70 78.46 10	17.60 22.27 10

Main effect means and n's:

Sequence	A mean=16.08, n=252; B mean=19.98, n-84
Presentation	M_{P1}=18.54,n=96; M_{P2}=17.08,n=120; M_{P3}=15.82,n=120
Introduction	M_{I1}=18.15, n=112; M_{I2}=17.92,n=112; M_{I3}=15.08,n=112

APPENDIX D1 (continued)

Frequency Distribution Immediate Post-Test--Open-End Times

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (S)	1	663.80	22.809*
Presentation (P)	2	54.92	1.887
Introduction (I)	2	163.96	5.634*
S x P	2	30.86	1.060
S x I	2	16.03	<1
P x I	4	18.06	<1
S x P x I	4	30.20	<1.037
Within cells	316[1)]	29.10	

1) See note on analysis of open-end scores.

Cell means(first row), variances(second row), and n's(bottom row):

A Group

	P1	P2	P3
I1	23.12 27.85 24	27.96 36.82 24	25.46 29.22 24
I2	25.57 21.08 30	27.67 27.54 30	25.37 26.52 30
I3	25.23 33.84 30	25.97 31.34 30	24.73 27.65 30

B Group

	P1	P2	P3
I1	20.12 19.27 8	22.25 18.50 8	20.00 44.86 8
I2	22.70 30.68 10	23.50 51.39 10	25.60 11.60 10
I3	20.10 14.77 10	24.10 14.54 10	22.60 53.38 10

Main effect means:

Sequence A mean = 25.69; B mean = 22.44

Presentation $M_{P1} = 24.33$; $M_{P2} = 25.63$; $M_{P3} = 24.55$

Introduction $M_{I1} = 23.82$; $M_{I2} = 26.20$; $M_{I3} = 24.61$

APPENDIX D1 (continued)

Frequency Distribution Immediate Post-Test--Multiple-Choice Scores

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (S)	1	1,077.16	8.040*
Presentation (P)	2	79.40	<1
Introduction (I)	2	362.59	2.714
S x P	2	165.66	1.244
S x I	2	112.52	<1
P x I	4	10.77	<1
S x P x I	4	181.97	1.358
Within cells	316[1]	133.97	

1) See note on analysis of open-end scores.

Cell means(first row), variances(second row), and n's(bottom row):

A Group

	P1	P2	P3
I1	33.08 113.38 24	29.17 110.41 24	29.29 165.09 24
I2	34.73 100.82 30	29.63 159.41 30	30.87 111.77 30
I3	30.33 124.44 30	30.63 173.48 30	28.03 106.52 30

B Group

	P1	P2	P3
I1	39.13 128.70 8	37.75 101.64 8	33.50 35.14 8
I2	31.30 180.46 10	39.00 123.78 10	28.20 200.62 10
I3	38.60 131.38 10	33.20 268.40 10	33.60 97.16 10

Main effect means:

Sequence	A mean = 30.65; B mean = 34.79
Presentation	$M_{P1} = 32.08$; $M_{P2} = 32.01$; $M_{P3} = 31.03$
Introduction	$M_{I1} = 33.55$; $M_{I2} = 31.54$; $M_{I3} = 29.96$

APPENDIX D1 (continued)

Frequency Distribution Immediate Post-Test--Multiple-Choice Times

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (S)	1	92.28	4.571*
Presentation (P)	2	13.30	<1
Introduction (I)	2	147.49	7.368*
S x P	2	8.14	<1
S x I	2	40.98	2.047
P x I	4	17.10	<1
S x P x I	4	5.81	<1
Within cells	316[1]	20.02	

[1] See note on analysis of open-end scores.

Cell means(first row), variances(second row), and n's(bottom row):

A Group

	P1	P2	P3
I1	28.46 15.74 24	25.67 23.19 24	27.21 23.13 24
I2	27.90 23.54 30	24.50 22.81 30	27.53 11.98 30
I3	27.77 20.81 30	25.80 25.68 30	26.80 9.54 30

B Group

	P1	P2	P3
I1	26.38 13.41 8	27.12 37.55 8	24.88 30.41 8
I2	25.80 17.73 10	23.50 29.17 10	25.00 28.67 10
I3	27.60 15.82 10	26.40 9.82 10	24.20 11.29 10

Main effect means:

Sequence A mean = 26.83; B mean = 25.62

Presentation M_{P1} = 26.86; M_{P2} = 26.18; M_{P3} = 26.61

Introduction M_{I1} = 27.66; M_{I2} = 25.34; M_{I3} = 26.55

APPENDIX D2

Central Tendency--Immediate Post-Test, Open-End Scores

Summary table: Analysis of Variance

Source	df	mean square	F
Sequence (AlvsB)	1	181.42	2.816
Presentation (P)	2	189.99	2.948
Introduction (I)	2	47.00	<1
S x P	2	384.32	5.966*
S x I	2	37.47	<1
P x I	4	36.28	<1
S x P x I	4	216.40	3.359*
Within cells	159	64.42	

Cell means(first row), variances(second row), and n's(bottom row):

A1 Group

	I1	I2	I3
P1	29.50 172.30 6	22.00 122.00 6	23.00 117.20 6
P2	17.00 72.50 5	11.40 65.30 5	24.80 23.20 5
P3	15.75 68.25 4	23.75 48.25 4	20.75 28.25 4

B Group

	I1	I2	I3
P1	23.11 67.28 18	20.44 87.08 18	24.61 61.55 18
P2	26.27 28.07 15	28.73 36.50 15	24.80 14.31 15
P3	23.42 84.45 12	19.42 97.54 12	19.75 30.02 12

Main effect means and n's:

Sequence A1 mean=21.20, n=45; B mean=23.52, n=135

Presentation M_{P1}=23.25,n=72; M_{P2}=24.38,n=60; M_{P3}=20.67,n=48

Introduction (n=60 for each) M_{I1}= 23.60; M_{I2}= 21.93; M_{I3}= 23.28

APPENDIX D2 (continued)

Central Tendency--Immediate Post-Test, Open-End Times

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (AlvsB)	1	145.19	7.589*
Presentation (P)	2	47.61	2.488
Introduction (I)	2	7.34	<1
S x P	2	3.35	<1
S x I	2	15.34	<1
P x I	4	3.31	<1
S x P x I	4	5.69	<1
Within cells	159	19.31	

Cell means(first row), variances(second row), and n's(bottom row):

A1 Group

	I1	I2	I3
P1	13.83 7.77 6	17.17 17.37 6	15.17 8.57 6
P2	14.20 1.70 5	16.60 19.30 5	17.00 6.50 5
P3	13.75 8.25 4	13.50 .33 4	13.25 8.25 4

B Group

	I1	I2	I3
P1	18.11 32.81 18	18.06 15.82 18	16.78 10.89 18
P2	17.67 37.10 15	17.47 13.55 15	17.07 23.50 15
P3	15.83 9.97 12	16.17 12.88 12	16.17 37.42 12

Main effect means and n's:

Sequence	A1 mean=15.07, n=45; B mean=17.14, n=135
Presentation	M_{P1}=17.08,n=72; M_{P2}=17.03,n=60; M_{P3}=15.42,n=48
Introduction (all n's = 60)	M_{I1}= 16.50; M_{I2}= 17.02; M_{I3}= 16.35

APPENDIX D2 (continued)

Central Tendency--Immediate Post-Test, Multiple-Choice Scores

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (AlvsB)	1	31.30	1.396
Presentation (P)	2	27.75	1.238
Introduction (I)	2	14.55	<1
S x P	2	33.58	1.498
S x I	2	7.72	<1
P x I	4	18.07	<1
S x P x I	4	33.95	1.515
Within cells	159	22.41	

Cell means(first row), variances(second row), and n's(bottom row):

A1 Group

	I1	I2	I3
P1	17.17 18.17 6	12.33 19.87 6	13.67 12.67 6
P2	12.20 53.70 5	12.60 6.30 5	14.40 6.80 5
P3	12.00 20.67 4	16.00 2.67 4	12.50 20.33 4

B Group

	I1	I2	I3
P1	13.67 26.47 18	13.50 29.32 18	13.72 27.74 18
P2	15.00 14.14 15	17.13 12.41 15	13.07 13.93 15
P3	16.00 32.73 12	15.75 42.75 12	15.50 8.64 12

Main effect means and n's:

Sequence	A1 mean=13.71, n=45; B mean=14.67, n=135
Presentation	M_{P1}=13.82,n=72; M_{P2}=14.57,n=60; M_{P3}=15.19,n=48
Introduction (all n's = 60)	M_{I1}= 14.58; M_{I2}= 14.83; M_{I3}= 13.88

APPENDIX D2 (continued)

Central Tendency--Immediate Post-Test, Multiple-Choice Times

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (AlvsB)	1	508.47	20.070*
Presentation (P)	2	101.06	3.990*
Introduction (I)	2	35.17	1.388
S x P	2	242.60	9.576*
S x I	2	13.01	<1
P x I	4	19.13	<1
S x P x I	4	27.67	1.092
Within cells	159	25.335	

Cell means(first row), variances(second row), and n's(bottom row):

Al Group

	I1	I2	I3
P1	33.33	29.67	32.83
	40.27	37.87	18.97
	6	6	6
P2	24.80	23.00	29.20
	16.20	10.50	58.70
	5	5	5
P3	26.00	30.25	29.25
	46.00	14.92	37.58
	4	4	4

B Group

	I1	I2	I3
P1	30.67	30.06	32.83
	27.41	18.64	26.97
	18	18	18
P2	32.53	32.73	33.13
	29.55	19.50	20.12
	15	15	15
P3	36.00	35.42	34.42
	10.18	16.81	36.27
	12	12	12

Main effect means and n's:

Sequence	Al mean=28.93, n=45; B mean=32.81, n=135
Presentation	M_{P1}=31.38,n=72; M_{P2}=31.01,n=60; M_{P3}=33.58,n=48
Introduction (all n's = 60)	M_{I1}= 31.67; M_{I2}=31.18; M_{I3}=32.68

APPENDIX D3

Frequency Distribution Retention Test--Open-End Scores

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (AlvsA2)	1	1200.87	13.543*
Presentation (P)	2	111.76	1.260
Introduction (I)	2	18.59	<1
S x P	2	44.33	<1
S x I	2	87.06	<1
P x I	4	217.43	2.452*
S x P x I	4	112.14	1.265
Within cells	96	86.67	

Cell means(first row), variances(second row), and n's(bottom row):

A1 Group

	P1	P2	P3
I1	29.33 49.47 6	19.33 31.47 6	26.83 30.17 6
I2	15.00 194.67 4	13.50 203.00 4	27.75 27.58 4
I3	22.11 74.86 9	26.00 227.00 9	18.56 84.28 9

A2 Group

	P1	P2	P3
I1	14.17 24.57 6	17.50 119.50 6	18.00 69.20 6
I2	14.50 135.00 4	16.50 33.67 4	14.25 86.25 4
I3	18.56 130.28 9	18.67 75.75 9	10.44 14.03 9

Cell means and n's of the significant P x I interaction

P1 groups (n = 12)	$M_{I1} = 21.75$	$M_{I2} = 18.42$	$M_{I3} = 22.42$
P2 groups (n = 8)	$M_{I1} = 14.75$	$M_{I2} = 15.00$	$M_{I3} = 21.00$
P3 groups (n = 18)	$M_{I1} = 20.33$	$M_{I2} = 22.33$	$M_{I3} = 14.50$

Main effect means and n's:

Sequence:	A1 mean=22.42, n=57; A2 mean=15.93, n=57
Presentation:	M_{P1}=20.86,n=36; M_{P2}=16.92,n=24; M_{P3}=19.06,n=54
Introduction: (all n's = 38)	M_{I1}=19.61; M_{I2}=19.55; M_{I3}=18.37

APPENDIX D3

Frequency Distribution Retention Test--Multiple-Choice Scores

Summary Table: Analysis of Variance

Source	df	mean square	F
Sequence (A1vsA2)	1	1038.04	7.001*
Presentation (P)	2	71.40	<1
Introduction (I)	2	8.32	<1
S x P	2	57.48	<1
S x I	2	523.16	3.529*
P x I	4	264.48	1.784
S x P x I	4	159.98	1.079
Within cells	96	148.26	

Cell means(first row), variances(second row), and n's(bottom row):

A1 Group,

	P1	P2	P3
I1	44.67 259.07 6	28.17 124.17 6	37.33 36.27 6
I2	29.50 169.67 4	27.75 411.58 4	43.25 74.92 4
I3	35.22 122.69 9	38.89 253.86 9	42.11 127.86 9

A2 Group

	P1	P2	P3
I1	28.17 46.17 6	33.50 298.70 6	28.17 44.97 6
I2	24.75 282.92 4	33.25 24.92 4	36.00 66.67 4
I3	33.44 106.78 9	36.78 164.69 9	24.22 110.94 9

Cell means of the significant S x I interaction (n=19 for each):

A1 groups $M_{I1} = 37.00$ $M_{I2} = 33.16$ $M_{I3} = 40.84$

A2 groups $M_{I1} = 29.95$ $M_{I2} = 35.00$ $M_{I3} = 27.95$

Main effect means and n's:

Sequence: A1 mean = 37.00, n = 57; A2 mean = 30.96, n = 57

Presentation: $M_{P1} = 33.33, n=36$; $M_{P2}=32.42, n=24$; $M_{P3}=35.11, n=54$

Introduction: $M_{I1}=33.47$; $M_{I2}=34.08$; $M_{I3}=34.39$
(all n's = 38)

Appendix D4

4) Pearson Product Moment Correlations Between Variables
Statistics Experiment, Fall, 1963

	Immediate Test Scores			Retention Test Scores		Times			
Variable number	2	3	4	5	6	7	8	9	10
A1 Group (n=32)	FDMC	CTOE	CTMC	ROE	RMC	FDOE	FDMC	CTOE	CTMC
1.FDI_Open End Sc.	.63	.70	.57	.75	.75	.06	.04	-.14	.13
2.FDI_M C Score		.53	.58	.62	.72	-.17	.36	-.42	.09
3.C T -Open End Sc.			.70	.73	.64	.19	.15	-.09	.34
4.6 T -M C Score				.67	.69	.11	.10	-.10	.34
5.FD- Retention OE					.75	.08	.20	-.06	.44
6.FD-Retention MC						-.20	.35	-.19	.19
7.FDI-OE Time							-.70	.20	.10
8.FDI_MC Time								-.05	.22
9.CT-OE Time									.12
B Group (n=105)									
1. FLI-Open End Sc.	.70	.56	.41			.00	.12	.12	.14
2. FDI_MC		.59	.47			-.13	.14	-.04	.18
3.CT-Open End Sc.			.62			-.05	.10	.01	.17
4.CT_MC Score						-.03	.06	-.15	.23
7.FDI-OE Time							.07	.53	.10
8.FDI_MC Time								.05	.45
9. CT-OE Time									-.21
A2 Group (n=77)									
1.FDI-Open End Sc.	.70			.70	.69	-.04	.26		
2.FDI_MC Score				.71	.76	-.06	.16		
5.FD-Retention-OE					.78	.08	.09		
6. FD-Retention-MC						-.07	.15		
7.FDI-OE Time							-.57		